

SEMANTICS AND MORPHOLOGY:

The Acquisition of Grammatical Gender in Russian

Yulia Rodina



A thesis submitted to the Faculty of Humanities,
University of Tromsø, November 2007

Acknowledgments

Writing this dissertation was both fun and hard work. Although fun part was first, I will start with hard work. This would be much harder without my supervisors, Marit Westergaard and Tore Nettet. I feel privileged to have had them both as my dissertation advisers. Their enthusiasm and commitment made me feel part of a team working on a joint project. I must thank them for the brainstorming discussions, for their input on the earlier versions of the manuscript, for their comments and suggestions, and for always being available.

I wish to thank Marit for introducing me to the world of language acquisition research. Starting with the Master course “Language Acquisition” I became interested in this type of research thanks to Marit’s enthusiasm, her knowledge, and her reasonable approach to the theories. Her expertise in the field truly increased the level of the research and her critical mind helped improving the dissertation in innumerable ways. Thank you, Marit, for getting deeply into the Russian gender system.

I would also like to thank Tore for his commitment to supervising this project from the very start and helping me till the very last moment. I thank him for sharing his knowledge about gender and about Russian in general with me. I am indebted to him for my intellectual growth, for structuring my thoughts and for keeping me focused. His advice and support academically and beyond have been invaluable.

I have also benefited from fruitful discussions with Bruce Morén, Tanja Kupisch, Tom Roeper. I thank Harald Clahsen for showing me where to start with the empirical part. Many thanks to Greville Corbett, Östen Dahl, Hans Olaf Enger, Donald Steinmetz and other participants of the Amsterdam Gender Colloquium for the useful advice and feedback.

I thank the little ones for making this work a fun experience. I also thank the teachers at the day-care centers in Ivanovo for allowing me to carry out the experiments and the parents for agreeing to read the silly books with me. Many thanks to two Russian families in Tromsø for the possibility to test the design and for their hospitality.

I greatly acknowledge the financial support of CASTL and the Department of Humanities that allowed me to carry out the experimentation. A special thanks to CASTL and its director Curt Rice for creating an outstanding linguistic environment, for the opportunity to visit international conferences and meet great researchers in Tromsø and outside.

I very much enjoyed my time as a PhD student at the University of Tromsø and I would like to thank all colleagues and friends for making life inside and outside linguistics so great. Thanks especially to Merete Anderssen, Kristine Bentzen, Bruce Morén, Sylvia Blaho, Peter Muriungi, Björn Lundquist, Marleen Van De Vate, Marina Pantcheva, Pavel Caha, Peter Jurgec, and Christine Østbø.

I wish to thank my very true friends Zhenya Romanova and Marina Dyakonova who I grew up together as a linguist for their support and encouragement. I miss you girls. Thanks especially to my dear friends in Tromsø and other places in the world: Monika Basic, Lucius Bader, Kaori Takamine, Tom Hansen, Patrycja Jablonska, Svetoslav Marinov, Jonny Jamtjord, Erlend Lien, and Madeleine Halmøy.

I want to thank my husband, Adnan Icagic, who helped me in so many ways: providing me with the literature when I started the project in exile, drawing pictures and making books for children, and always believing in me. Most of all I wish to thank him for taking care of me in the last days of panic, for his patience and understanding.

Finally, I am indebted to my parents, my sister and my grandmother Nadezhda for their moral support and their love.

Contents

Acknowledgments	iii
Abbreviations	ix
1 Introduction	1
1.1 Overview	1
1.2 Structure of the dissertation	4
1.3 Key concepts	5
2 The gender system of Russian: Criteria and relationships	7
2.1 Introduction	7
2.2 The semantic criterion	9
2.3 The morphological criterion	11
2.4 Agreement	14
2.4.1 <i>Papa</i> -type nouns and male names in <i>-a</i>	16
2.4.2 Double gender nouns	16
2.4.3 Hybrids	19
2.4.4 Female names in <i>-ok/-ik</i>	22
2.4.5 Summary	23
2.5 Semantic hierarchy vs. default hierarchy	23
2.6 Summary of the chapter	28
3 Previous studies on the acquisition of grammatical gender	31
3.1 Introduction	31
3.2 Semantic and formal criteria in acquisition	33
3.2.1 French	35
3.2.2 German	36
3.2.3 Hebrew	38
3.2.4 Czech	39
3.2.5 Summary	40
3.3 Gender acquisition in Russian	41
3.3.1 Gender acquisition from onset till the age of 3	41
3.3.2 Summary of previous research on Russian	52

3.4	Previous research on children's acquisition of variation in the input	53
3.5	Summary of the chapter	55
4	General predictions and research questions	57
4.1	Introduction	57
4.2	Frequency issues	59
4.2.1	The words and rules model	60
4.2.2	The rules and competition model	61
4.2.3	Rule- vs. rote-based gender acquisition?	62
4.3	Acquisition of variable forms	64
4.4	Referential gender	65
4.5	Classes of nouns	65
4.6	Summary	66
5	Methodology	67
5.1	Introduction	67
5.2	Description of the participants	67
5.3	Elicited production	68
5.4	Data collection	69
5.5	Transcription	71
5.6	Experiment 1: <i>Papa</i> -type nouns	71
5.6.1	Goal of the experiment	71
5.6.2	Materials and procedure	72
5.7	Experiment 2: Male names in <i>-a</i>	73
5.7.1	Goal of the experiment	73
5.7.2	Materials and procedure	74
5.8	Experiment 3: Novel nouns	75
5.8.1	Goal of the experiment	75
5.8.2	Materials and procedure	76
5.9	Experiment 4: Female names in <i>-ok/-ik</i>	77
5.9.1	Goal of the experiment	77
5.9.2	Materials and procedure	78
5.10	Experiment 5: Hybrid nouns - referent absent	79
5.10.1	Goal of the experiment	79
5.10.2	Materials and procedure	79
5.11	Experiment 6: Hybrid nouns - referent present	80
5.11.1	Goal of the experiment	80
5.11.2	Materials and procedure	80
5.12	Double gender nouns	82
5.12.1	Goal of the experiment	82
5.12.2	Materials and procedure	82
5.13	Double gender nouns	83
5.13.1	Goal of the experiment	83
5.13.2	Materials and procedure	84

5.14	Summary of the chapter	85
6	Acquiring gender in obligatory contexts	87
6.1	Introduction	87
6.2	Hypotheses and predictions	88
6.3	Results	92
6.3.1	Results of the main study	92
6.3.2	Results of the pilot test	98
6.3.3	Summary	99
6.4	Overregularization in gender acquisition	100
6.4.1	Age 3;0	100
6.4.2	The (non-)effect of frequency on the acquisition of the semantic rule	102
6.4.3	Proper names vs. common nouns in gender acquisition	106
6.4.4	Blocking as an explanation for the overregularization and the course of gender acquisition	110
6.5	Summary and conclusion	112
7	Acquiring gender in variable contexts	115
7.1	Introduction	115
7.2	Hypotheses and predictions	117
7.3	Results	118
7.4	Obligatory vs. Variable Contexts	125
7.5	child vs. caregiver data	127
7.6	development	131
7.7	Summary and conclusion	135
8	Referential gender	139
8.1	Introduction	139
8.2	Hypotheses and predictions	140
8.3	Results	143
8.3.1	Double gender nouns - referent present	143
8.3.2	Double gender nouns vs. hybrids - referent absent	149
8.3.3	Summary of the results	151
8.4	Knowledge of referential gender	152
8.5	Summary of the chapter	157
9	Wrapping things up	159
9.1	Introduction	159
9.2	Overview	160
9.3	The course of acquisition	162
9.3.1	The general developmental pattern	162
9.3.2	Cue-based gender acquisition	164
9.3.3	Non-simultaneous acquisition of the semantic criterion	168

9.4	Gender assignment and gender learnability	173
9.5	Innateness and learning	176
10	Summary and conclusions	181
10.1	Summary of individual chapters	181
10.2	Ideas for further research	184
10.3	Concluding remarks	185
	Appendix I	189
	Appendix II	191
	Appendix III	209

Abbreviations

ACC	accusative
ADJ	adjective
ARG	agreement
ATTRIB	attributive
DAT	dative
DIM	diminutive
F	feminine
GEN	genitive
INS	instrumental
LOC	locative
M	masculine
N	neuter
N	number (in Chapters 6-8)
NOM	nominative
NP	noun phrase
PL	plural
poss	possessive
Prn	pronoun
PST	past
RC	the Rules and Competition model
refl	reflexive
SG	singular
UG	Universal Grammar
V	verb
WR	the Words and Rules model

Chapter 1

Introduction

1.1 Overview

The study of gender as an agreement class in language can shed light on important aspects of human categorization and interaction of categories in the human mind. The goal of this dissertation is to present experimental research that provides a deeper understanding of the role of natural gender in a grammatical system of a language on the basis of evidence from the first language acquisition of Russian.

For children learning their first language, as for adults learning their second language, gender can be viewed as a challenging domain, “since it requires a complex and often arbitrary integration of both formal (syntactic and phonological) and general cognitive factors related to meaning and reference, as well as wide cross-linguistic variation” (Lust 2006:250). Nevertheless, in the previous century researchers were surprised to find that very young children (under the age of 2) were able to derive a noun’s gender from language-specific, highly abstract, formal criteria and quickly develop a system of formal gender assignment rules (Karmiloff-Smith 1979, Clark 1985, Mills 1986, Henzl 1975, Smoczyńska 1985, Levy 1983a;b, Gvozdev 1961, Popova 1973 *inter alia*). At the same time semantic gender criteria seemed to be beyond the scope of young children. The establishment of formal gender regularities by young children has been explored in various languages (e.g. French, German, Czech, Hebrew, Russian, etc.), yet little is known about how the integration of the semantic criteria proceeds.

This dissertation investigates exceptional classes of nouns, whose gender in the adult language is derived from the semantic (sex-based) criterion, yet their form is contradictory. The first class in (1-a) illustrates masculine nouns in the second declension and includes two subclasses: male kinship terms like

papa ‘daddy’ (henceforth *papa*-type nouns) and male names in *-a* like *Vanya* derived from full names like *Ivan*. Both subclasses have morphology typical of feminine nouns, yet they denote males. The second class presented in (1-b) consists of two subclasses as well: hybrids which denote various professional titles, e.g. *doktor* ‘doctor’, and female names in *-ok/-ik*, e.g. *Ninčik*, derived by masculine suffixes from female names like *Nina*. The third class in (1-c) are so-called common gender or double gender nouns that typically characterize individuals by some personal trait (often negative) or behavior.

- (1) a. (i) male kinship terms (e.g. *papa* ‘daddy’, henceforth *papa*-type nouns);
 - (ii) male names in *-a* (e.g. *Vanya* from full name *Ivan*);
- b. (i) hybrids (e.g. *doktor* ‘doctor’);
 - (ii) female names in *-ok/-ik* (e.g. *Lizok* from *Liza*);
- c. common gender nouns or double gender nouns (e.g. *plaksa* ‘cry-baby’).

A detailed discussion of the classes of nouns in (1) will be given in Chapter 2. For now it is sufficient to notice that the gender of hybrids and double gender can vary depending on the referent’s biological sex, yet their form stays the same. Even more variation is attested for the subtypes listed in (1-b), which generate intricate agreement patterns. Henceforth, I will refer to them as nouns used in variable contexts, while the subtypes in (1-a) and (1-c) will be called nouns used in obligatory contexts.

The main purpose of this study is to investigate how the discovery of gender proceeds in the mind of a language learner. The central topic of the discussion is the role of the semantic criterion in gender acquisition, more specifically, its integration into the grammatical system of a language and its relationship with formal gender criterion. This research is particularly concerned with the establishment of the semantic rule across various noun classes and the question whether this process is simultaneous or not. These issues lead to the following general research questions:

- (2) a. When and how is the semantic rule acquired for various classes of nouns?
- b. What factors underlie the process of acquisition?

In order to account for how gender knowledge grows in children, this study explores a number of issues which have received considerable attention in the acquisition research in the past few years. One of them is the role of input frequencies in gender acquisition. The rule- vs. rote-based nature of the semantic principle is explored through the prism of children’s overreg-

ularization errors for individual nouns. The findings are discussed in light of two theories of morphological acquisition: the Words and Rules model (Pinker 1999) vs. the Rules and Competition model (Yang 2002). I show that gender is a domain sensitive to frequency in the sense of Pinker’s dual-mechanism model. Yet, based on evidence from the acquisition of proper names I argue that frequency is not the only factor that constrains children’s overregularization tendencies. I suggest that the differences in the semantic representation of common nouns vs. proper names may account for the asymmetries observed in children’s production.

Another important issue which I address in the present study is children’s acquisition of variation in the input. As pointed out by Roberts (1997a:354), “a complete acquisition model demands the inclusion of all forms of language, those which are variable and those which are categorical”. The comparison of children’s behavior for nouns in obligatory vs. variable contexts as well as children’s vs. caregivers’ behavior in variable contexts will reveal a delay in the acquisition of the semantic rule for nouns in variable contexts. This result is taken to indicate that children may lack the socio-linguistic competence and that gender is a grammatical category for them and not a socio-linguistic phenomenon. In the final discussion in Chapter 9, this delay as well as other asymmetries in the acquisition of the semantic rule across the various noun classes are accounted for in terms of qualitative and quantitative properties of the input.

An important role in my investigation is reserved to the notion of referential gender. Specifically, I investigate children’s ability to assign gender to a noun via identification with a human referent in a concrete discourse situation. I show that children treat double gender nouns as a class different from hybrids. It is thus established that children are sensitive to classes of nouns. Thus this dissertation provides converging evidence for the claim that children are distinguish classes of categories (Roeper 2007).

Among other important findings, the examination of the course of acquisition of gender in Russian reveals a similar pattern observed in various other languages. Based on the novel empirical evidence presented in this dissertation, I show that gender acquisition proceeds from formal to semantic, i.e. formal grammatical analyses are gradually replaced by those based on semantic (sex-based) distinctions. As my own contribution to the field, in this dissertation I make an attempt to account for the factors that affect this process and the mechanisms that underlie it. Based on the acquisition pattern for various subtypes of nouns that emerges from my data I argue for a cue-based approach to gender acquisition in Russian (Lightfoot 1999, Westergaard 2006). I suggest that there are separate semantic cues for the individual noun classes distinguished by children, which may be seen as formal

representations of very specific semantic rules that children apply “locally” to each subtype of nouns in particular.

I finally show that children’s agreement behavior with various subtypes of nouns may offer an insight into how gender knowledge is organized in the mental grammar of a child. In particular, I address the issues of learnability and innateness and argue against the core function of the semantic principle (Corbett 1991). I also develop the idea that gender acquisition may be seen as an innately guided learning process.

1.2 Structure of the dissertation

Chapter 2 outlines the Russian gender system with regard to the classes of nouns under investigation. The discussion focuses on the following relevant factors: 1) the semantic criterion (i.e. biological sex); 2) the morphological criterion (i.e. declension class); and 3) agreement patterns. In the second part of this chapter the relationship of the semantic and morphological gender criteria are discussed in terms of two gender assignment theories: the assignment rules hierarchy (Corbett 1991) and the default (or markedness) hierarchy (Rice 2006). Chapter 2 ends with the research questions with regard to child first language acquisition that emerge out of the issues presented in that chapter.

In Chapter 3, some previous studies on the acquisition of gender in Russian as well as in some other languages are reviewed, with reference to their relevance for the study of the relationships between formal and semantic gender criteria. This chapter also focuses on the issue of input variability and considers some relevant findings from previous research on the acquisition of variable input.

In Chapter 4 I present the general research questions and predictions based on the previous acquisition studies as well as certain theoretical considerations that follow from the gender assignment theory considered in Chapter 2. In addition this chapter focuses on the issue of input frequencies and reviews two theories of morphological acquisition: the Words and Rules model (Pinker 1999) vs. the Rules and Competition model (Yang 2002). It also shows how their assumptions can be generalized to the acquisition of gender.

This dissertation takes an experimental approach to its discussion matter. Using an Elicited Production Technique, agreement forms are elicited from monolingual Russian children (age 2;6 - 4;0), and in certain tasks from their caregivers’ as well as from 12 older children (age 5;1-6;5). Chapter 5 presents the methodology used in the present study. This chapter provides the description of the participants, method, and design of eight experimental

tasks. It also includes the details of data collection and transcription.

The results of the experimentation are presented and discussed in three chapters. Chapter 6 contains an analysis of the acquisition of the semantic rule for nouns in obligatory contexts, i.e. for *papa*-type nouns and male names in *-a*. Chapter 7 focuses on children's behavior in variable contexts, i.e. for hybrids and female names in *-ok/-ik*. Chapter 8 investigates children's knowledge of referential gender with regard to double gender nouns.

Chapter 9 summarizes the results of the study and provides a comprehensive analysis of the gender acquisition process. Finally, Chapter 10 contains a summary, ideas for further research, and some concluding remarks.

1.3 Key concepts

The definitions below are crucial for understanding the linguistic phenomenon examined in the current study. They are partly taken from Corbett (1991), Aronoff (1994), Aronoff and Fudeman (2005) and compiled into a coherent whole here.

- (3) **Grammatical gender** A grouping of nouns in a language on the basis of semantics, morphology, phonological shape, arbitrary characteristics, or a combination thereof. Gender is also called 'noun class' in scholarly literature. Importantly, in this dissertation, I use the term 'gender' to refer to masculine, feminine and neuter distinctions and the term 'noun class' to refer to types or subcategories of nouns.
- (4) **Gender assignment** The operation of attributing a noun to an agreement class on the basis of its inherent, invariable characteristics. In the current study these are morphological and semantic (sex-based) criteria.
- (5) **Declension (Inflectional class)** is a set of lexemes whose members each select the same set of inflectional realizations. Russian has four main noun paradigms each consisting of 12 case forms (six in the singular and six in plural).
- (6) **Agreement (Concord)** is the process by which one lexical category is inflected to express the properties of another, e.g. a verb bearing number and gender morphology that reflect those of a subject.
- (7) **Semantic agreement** Agreement consistent with the gender assigned on the basis of semantic properties of a noun.
- (8) **Syntactic agreement (Grammatical agreement)** Agreement consistent with the gender assigned on the basis of morphological or

phonological properties of a noun.

- (9) **Agreement target** A grammatical category which shows agreement with the noun (agreement **controller**). The agreement targets considered in this study are adjectives and verbs in the past tense.

Chapter 2

The gender system of Russian: Criteria and relationships

2.1 Introduction

This chapter considers the gender system of Russian with a special focus on the nouns I investigate in this dissertation. Russian is known as a three-way gender system, where nouns belong to a masculine, feminine or neuter gender. This distinction is made on the basis of agreement, i.e. “[t]he process by which one lexical category (e.g. an adjective) is inflected to express the properties of another (i.e. a noun). . .” (Aronoff and Fudeman 2005:234). Therefore, since Russian nouns occur in three distinct sets of syntactic frames, they are divided into three genders: masculine, feminine, and neuter, as shown in (1).^{1,2,3}

- (1) a. Strann-yj junoša ležal-Ø na skamejke.
strange_M youth_(M) lie_{PST.M} on bench
‘A strange youth was lying on the bench.’
b. Strann-aja devuška ležal-a na skamejke.
strange_F girl_(F) lie_{PST.F} on bench
‘A strange girl was lying on the bench.’

¹In the glosses, the gender of a noun is marked in parentheses, and the gender of an agreeing item (target) is marked without. The nouns are shown in nominative singular, unless stated otherwise.

²Each of the three genders is divided by Corbett (1991) into two subgenders: animate and inanimate. More specifically, animate nouns show accusative-genitive syncretism, which is not observed for inanimates. Since this issue is not particularly relevant for the discussion here, it is not considered further.

³In (1-c) the symbol ‘ transliterates the Russian soft sign, which indicates palatalization of the preceding consonant.

- c. Strann-oe pis'mo ležal-o na skamejke.
 strange_N letter_(N) lie_{PST.N} on bench
 'A strange letter was lying on the bench.'

Among the agreement targets which reflect the gender of Russian nouns are adjectives (long and short), participles, verbs, pronouns (personal, relative, possessive, and demonstrative), and numerals (cardinal (*odin* 'one' and *dva* 'two') and ordinal). Adjectives and verbs are considered to be canonical agreement targets (Corbett 2006). Note that adjectives can be used attributively as well as predicatively, verbs express agreement in the past tense only. Note also that gender distinctions are available only in the singular.

Starting with Hockett (1958:231), who defined genders as classes of nouns "reflected in the behavior of associated words", agreement has been considered as the determining criterion of gender. While in Russian there are three classes of nouns which are distinguished syntactically by agreement, there are other languages that have a more extensive gender system. For example, as reported in Aronoff and Fudeman (2005), Kujamaat Jóola, a language in the Niger-Congo family (Atlantic languages) has 19 noun classes, and Gombe Fula has 25. Yet, agreement being the gender indicator depends on phonological, morphological, and semantic properties of the noun. According to Corbett (1991), semantic criteria can include a number of options: 'male/female', 'animate', as well as some fine-grained distinctions e.g. 'insects', 'bones', etc. The gender of Russian nouns is defined on the basis of two factors, viz. semantic ('male/female') and morphological (declensional class), which rather often overlap, pointing towards the same gender. However, in some cases, which are the main target of this dissertation, the two factors can be contradictory.

It should be noted that in Russian natural gender can also be syntactically realized with first and second singular pronouns *ja* 'I' and *ty* 'you'. These pronouns have only one form, i.e. they are not inflected for gender. However, forms agreeing with them show gender agreement. In (2) and (3) the verb forms agreeing with these pronouns reflect the sex of the speaker and the addressee respectively, while the controller is virtually absent from the syntactic phrase (from Corbett 1991:128):

- (2) a. ja čital-Ø (male speaker)
 I read_{PST.M}
 ‘I was reading.’
- b. ja čital-a (female speaker)
 I read_{PST.F}
 ‘I was reading.’
- (3) a. ty čital-Ø (male addressee)
 you read_{PST.M}
 ‘You were reading.’
- b. ty čital-a (female addressee)
 you read_{PST.F}
 ‘You were reading.’

The aim of this chapter is to provide a detailed description of the important features that define gender in Russian. Corbett (1991) undertook a detailed study of Russian and it is his account that I follow here. Sections 2.2 and 2.3 consider the two gender relevant criteria - morphological and semantic - with particular emphasis on the nouns under investigation. These are listed in (4):

- (4) a. (i) male kinship terms (e.g. *papa* ‘daddy’, henceforth *papa*-type nouns);
 (ii) male names in *-a* (e.g. *Vanya* from full name *Ivan*);
- b. (i) hybrids (e.g. *doktor* ‘doctor’);
 (ii) female names in *-ok/-ik* (e.g. *Lizok* from *Liza*);
- c. common gender nouns or double gender nouns (e.g. *plaksa* ‘cry-baby’).

Section 2.4 demonstrates that both semantics and morphology exert great influence on agreement with these nouns. For some of them agreement is rather straightforward, while others reveal more complex patterns. Section 2.5 discusses the relationships between these features in terms of the gender assignment theory with regard to two main approaches: semantic hierarchy by Corbett (1991) and default hierarchy by Rice (2006). A summary of the chapter is provided in Section 2.6, where I also formulate some questions with regard to child first language acquisition.

2.2 The semantic criterion

The semantic criterion involves the male vs. female distinction (i.e. a natural gender distinction). Biological sex is a criterion applicable only to animates,

more specifically, to nouns that denote human beings as well as domesticated and higher animals. Corbett (1991) calls these nouns sex-differentiable, as the language distinguishes a form for males and another for females, e.g. *suprug* ‘husband’ and *supruga* ‘wife’, or *papa* ‘daddy’ and *mama* ‘mommy’. As Corbett points out, one can be confident that a noun denoting a male is masculine, and a noun denoting a female is feminine. Thus, since *papa*-type nouns and male names in *-a* denote only males, they are unambiguously masculine, while female names in *-ok/-ik* must be feminine, as they denote only females. However, the sex specification is not always a part of a lexical entry of a noun. The other two classes of nouns that I consider in this dissertation, i.e. hybrids and double gender nouns, are different. As pointed out by Crockett (1976) and Dahl (2000) among other researchers, these nouns are semantically unspecified for sex, i.e. they are applicable to persons of either sex and the sex of their referents can only be determined in a situational context. Note also that these nouns use only one word form for males and females, i.e. they are not sex-differentiable. If the meaning of e.g. *papa* ‘daddy’ can be defined as ‘male parent’, where sex is included into the denotation, it can hardly be said that sex is part of the meaning of *doktor* ‘doctor’ or *plaksa* ‘crybaby’. The former denotes a person of medical profession and the latter describes a person on the basis of a certain personal behavior. This means that hybrids and double gender nouns are semantically unspecified for sex, or ‘sex-neutral’ (cf. Dahl 2000:106), i.e. sex specification is not part of their meaning, but introduced by the specific individual. Thus, in some cases the semantic criteria are introduced by the noun itself, and in the others they are introduced by the referent of the noun. From this perspective, the nouns in my study can be divided into two groups: nouns that are inherently specified for gender, i.e. male kinship terms, male names in *-a*, and female names in *-ok/-ik*, and nouns that are not, i.e. hybrids and double gender nouns.

Dahl (2000) made an important distinction between gender as a property of a noun (as a lexical item), and gender as a property of a noun phrase (as an occurrence in a particular discourse situation). The former he calls *lexical* gender; it applies to nouns proper and holds independently of particular occasions of utterances. The latter is called *referential* gender; it is established on the basis of the sex of the referent of a noun phrase in a particular situation. In other words, lexical gender is decided once and for all, while referential gender is decided on each occasion when the noun is used. From this perspective, *plaksa* ‘crybaby’ is masculine when it refers to a male and it is feminine when the referent is a female. With regard to hybrids, Dahl suggests that e.g. *vrač* ‘physician’ has two genders: lexical, i.e. masculine, when it refers to males or females, and referential, i.e. feminine, when it refers to females. This is the result of a diachronic change, such that in the

presocialist society (i.e. before 1917) the majority of occupations in Russia were male-dominated, hence hybrids were only masculine. In the course of the 20th century, when women began to be employed in what were traditionally male-dominated occupations, the grammatical status of hybrids has changed in the direction of double-gender nouns for whom semantic information is expressed by the concrete referent rather than by the noun proper (cf. Švedova 1980, Lopatin 1989). Nevertheless, the distinction between hybrids and double gender nouns holds, since for the latter sex criterion is an obligatory factor (when the noun functions as a subject), but not for the former. This important point comes out clearly in Section 2.4, where I consider agreement manifestation for these nouns.

Alexiadou (2004), who explores a similar phenomenon in various languages (Spanish, Greek, Italian), uses a somewhat different terminology. In particular, based on the fact that certain nouns in all languages are inherently specified for gender in the lexicon and others do not contain gender specification as an intrinsic property, she proposes a distinction between two types of gender: fixed and variable. The former is considered to be part of the nouns' intrinsic features (semantic or formal), while the latter must be assigned via identification with a human referent.

To conclude, the nouns considered in this dissertation differ with respect to how gender-relevant semantic information is expressed. For some nouns, i.e. male kinship terms, male names in *-a*, and female names in *-ok/-ik*, the sex criterion is available on the noun proper, i.e. it is expressed in the lexical semantic content of a noun. On the other hand, for hybrids and double gender nouns it is available through the noun's referent. This means that it is necessary for the child learning the language to know not only the meaning of a noun but to be able to make a connection between the referent and the noun itself. In addition to the semantic features the nouns under study possess formal, i.e. morphological, features, which however do not overlap. On the contrary, the two criteria are in conflict, as I show in the next section.

2.3 The morphological criterion

As I said in the previous section, the semantic criteria are only relevant for nouns denoting human beings and certain animals; hence they only cover a relatively small proportion of nouns, while for the majority of nouns in the language there is a strong correlation between gender and inflectional class, i.e. declension. According to Corbett (1991), Russian nouns form four inflec-

tional paradigms,⁴ and in general their gender is marked morphologically in a fairly regular manner, so that nouns of declensional type I are masculine, nouns of declensional type II and III are feminine, and nouns of declensional type IV are neuter. This distinction is demonstrated in Table 2.1. Note that the words are shown only in singular, as there are no gender distinctions in the plural.

Table 2.1: Declensional classes in Russian

	I	II	III	IV
	<i>suprug</i> (M)	<i>supruga</i> (F)	<i>sol'</i> (F)	<i>lico</i> (N)
	‘husband’	‘wife’	‘salt’	‘face’
<i>SG</i>				
NOM	<i>suprug-∅</i>	<i>suprug-a</i>	<i>sol'-∅</i>	<i>lic-o</i>
ACC	<i>suprug-a</i>	<i>suprug-u</i>	<i>sol'-∅</i>	<i>lic-o</i>
GEN	<i>suprug-a</i>	<i>suprug-i</i>	<i>sol'-i</i>	<i>lic-a</i>
DAT	<i>suprug-u</i>	<i>suprug-e</i>	<i>sol'-i</i>	<i>lic-u</i>
INS	<i>suprug-om</i>	<i>suprug-oj</i>	<i>sol'-ju</i>	<i>lic-om</i>
LOC	<i>suprug-e</i>	<i>suprug-e</i>	<i>sol'-i</i>	<i>lic-e</i>

As pointed out by Corbett (1991), gender assignment crucially depends on the knowledge of declensional classes, i.e. the whole paradigm of inflectional affixes.⁵ This is to say that gender cannot be simply derived based on one case form. For example in Table 2.1 all four nouns have distinct phonological forms in nominative singular, which is considered to be the basic form of a Russian noun (cf. Corbett 1991:35). Specifically, the masculine noun *suprug* ‘husband’ in class I ends in a non-palatalized (hard) consonant and has zero ending (-∅), its feminine counterpart *supruga* ‘wife’ in class II ends in -a, the feminine noun *sol'* ‘salt’ in class III ends in a palatalized consonant and has zero ending, and the neuter noun *lico* ‘face’ in class IV ends in -o. Thus, it might appear that a noun’s phonological form in nominative singular can be sufficient to predict its gender. However, as Corbett notes, there are examples where reference to a single form is not sufficient. Specifically, nouns that end in a palatalized consonant and have zero ending in nominative singular can be either feminine, e.g. *sol'* ‘salt’, or masculine, e.g. *den'* ‘day’. Nevertheless, the former belong to Declension III and the latter to Declension I. Thus it

⁴In later work Corbett and Fraser (2000) introduced declensional class V which covers indeclinable nouns like *pal'to* ‘coat’. These nouns are not considered here, since they fall outside the scope of the morphological assignment rules, which account for gender with the nouns in Table 2.1 and which are considered further in Section 2.5.

⁵Corbett (1991) argues that gender is not a feature on the stem, since different gender values are represented by different sets of inflectional affixes.

appears that the form of nominative singular for these nouns is ambiguous, hence it is only knowing the whole inflectional paradigm that gender of these nouns can be successfully predicted.⁶

From an acquisition perspective, phonological information or, to be more specific, the sound shape of the nominative singular is even more ambiguous, since many neuter nouns that belong to Declension IV and end in an unstressed *-o* (a reduced vowel, typically schwa) are undistinguishable from feminine nouns in Declension II that end in an unstressed *-a*, since both of them are reduced vowels (typically schwas), compare e.g.: *mám*[ə] ‘mommy’ (feminine) vs. *zérkal*[ə] ‘mirror’ (neuter).⁷

Unlike neuter nouns and nouns ending in palatalized consonants, the phonological form of nominative singular is a reliable gender predictor for the nouns considered in my investigation. These nouns belong to declension I or II and end in *-a* (stressed or unstressed) or a hard consonant and \emptyset in nominative singular.⁸ This fact is important, since the nouns in my experimental study will be presented to children in the nominative singular form.

With regard to their morphological properties, male kinship terms, male

⁶Interestingly, according to Zakharova (1973), declensions of feminine and masculine nouns ending in palatalized consonants are acquired by children only toward the end of preschool age, i.e. by 6/7. At the same time Gvozdev (1961) reports that gender errors with these nouns occur in the speech of his son Zhenya up to the age of 7;9. These errors include the attribution of feminine nouns like *sol'* ‘salt’ to masculine gender and might be due to child’s orientation to the form of a noun in nominative singular, which, according to Zakharova (1973:283), is “the more firmly acquired” case form than the forms of the oblique cases.

⁷Gvozdev (1961) reports that stem-stressed neuter nouns remain problematic for his son until late in the preschool years.

⁸One may notice here that unstressed *-a* is also ambiguous, since it is identical with the unstressed *-o*. Hence feminine nouns can be confused with neuter. However, the acquisition evidence suggests that only neuter nouns are confused with feminine but not the other way around. This may be due to fact that the class of neuter nouns (among which there are very frequent words like e.g. *moloko* ‘milk’) is considerably smaller than the class of feminine or masculine nouns (cf. Corbett 1991:78). The latter represent the majority in the lexicon. Interestingly, there is evidence that in some Russian dialects (to the south-east of Moscow) stem-stressed neuter nouns can take the inflections of the second (“feminine”) declension: e.g. *ubirat' sénu*_{ACC(F)} vs. Standard Russian *ubirat' séno*_{ACC(N)} (from Kasatkin 2005:122). This can be an indication that these neuter nouns are becoming feminine. Furthermore, it can be suggested that this language change arises in the child language (the idea that originates to Lightfoot (1979)). Polinsky (in press) also mentions that American Russian heritage speakers whose dominant language is English and whose proficiency level of Russian is rather low develop a two-way gender system as they assimilate stem-stressed neuter nouns to feminine, by analogy with feminine nouns in *-a*.

names in *-a*, and double gender nouns belong to declension II, where the majority of feminine nouns are found. This means that *papa* ‘daddy’, *Vanya*, and *plaksa* ‘crybaby’ are morphologically equivalent to feminine nouns like *mama* ‘mommy’. This also means that these nouns should be feminine according to their morphology, but masculine according to their semantics (in the case of double gender nouns this holds only when they refer to males). Clearly, the nouns’ morphological properties do not correlate with their semantic properties. A similar conflict is observed for hybrids when they refer to females, and for female names in *-ok/-ik*. These nouns belong to declension I, hence they should be masculine according to their morphology.⁹ However, they should be feminine according to the semantics, since these nouns refer to females.

It appears that semantic and morphological criteria make conflicting predictions about the nouns’ gender. This means that the gender information that the child finds in the input is rather ambiguous. Yet, there is another source of information, namely agreement, which various researchers consider to be the determining criterion of gender, since it directly reflects the gender value of the noun (Hockett 1958, Corbett 1991 *inter alia*; cf. the discussion in Section 2.1). Agreement is considered in detail in the next section, and it turns out to be rather straightforward for male kinship terms, male names in *-a*, and double gender nouns, but varied and complex for hybrids and female names in *-ok/-ik*.

2.4 Agreement

Agreement is an essential part of the gender phenomenon, as it follows from Hockett’s (1958) definition presented in Section 2.1. Agreement reflects the features of the controller (i.e. a noun in our case), on a particular target, e.g. an adjective, by means of inflection. As I said in the introduction to this chapter, Russian gender agreement affects various targets: adjectives, participles, pronouns, verbs, and numerals. The inflectional markers of the targets that show agreement with masculine and feminine nouns in the nominative singular are summarized in Table 2.2.¹⁰

⁹Note that full female names, e.g. *Liza*, belong to declension II, however, when the affective suffixes *-ok/-ik* are attached to them, they receive the inflectional paradigm of declension I, e.g. *Lizk-a_{ACC}*, *Lizk-om_{INS}*, etc.

¹⁰The full version should include other cases, but, as I said before, nominative singular is the basic form, which has also been elicited in the experimental study.

Table 2.2: Agreement markers in nominative singular for masculine and feminine genders

Gender	Target		
	Short adj. Verb Cardinal numeral Personal prn. Possessive prn.	Long adj. Participle Ordinal numeral Relative prn.	Demonstrative prn.
Masculine	-Ø <i>mamin-Ø</i> 'mommy _{poss.M} '	-yj/-oj <i>starš-ij</i> 'older _M '	-ot <i>et-ot</i> 'this _M '
Feminine	-a <i>mamin-a</i> 'mommy _{poss.F} '	-aja/-ija <i>starš-aja</i> 'older _F '	-a <i>et-a</i> 'this _F '

According to Corbett (2006), in agreement the feature specification of the target should match that of the controller. Consider, for example, the structure in (5), where feminine gender is found on the the adjective *novaja* 'new' as a consequence of its presence in the noun *kniga* 'book' (feminine). In other words, the target *novaja* is inflected to express the properties of the controller *kniga* 'book'.

- (5) nov-aja kniga
 new_F book_(F)
 'new book'

As we already know a noun can have two gender relevant features, viz. morphological and semantic (available only for nouns denoting (or referring to) human beings). From this perspective a more specific observation can be made about the example in (5), namely, the target *novaja* 'new' expresses the morphological feature of the controller *kniga* 'book'. Since *kniga* 'book' is inanimate, it is clearly the morphological criterion that exerts influence on agreement with this noun. However, in cases where the two features overlap, the situation is not straightforward. For example, the structure in (6) shows feminine agreement with the noun *mama* 'mommy' (feminine), which belongs to declension II and denotes a female. Since both semantics and morphology suggest the same gender, i.e. feminine, it is not clear which of them determines the agreement in this case.

- (6) naš-a mama
 our_F mommy_(F)
 ‘our mommy’

The nouns, which are the target of this dissertation, present examples of form-meaning mismatches, therefore, since the two gender features do not overlap, it can be clearly seen which one of them exerts influence on agreement. In fact, as I show in Sections 2.4.1 to 2.4.4, it is more often that semantics affects agreement than morphology.

2.4.1 *Papa*-type nouns and male names in *-a*

Papa-type nouns and male names in *-a* exhibit a very straightforward agreement pattern. For these nouns, which have feminine morphology but denote males, agreement is always masculine, hence it always has semantic justification. In other words, it is the semantic feature of the noun (i.e. ‘male’) which is reflected through agreement, as shown in (7).

- (7) a. Moj-Ø papa prišel-Ø.
 my_M daddy_(M) come_{PST.M}
 ‘My daddy came.’
 b. Vanya byl-Ø sovsem zdorov-Ø.
 Vanya_(M) be_{PST.M} absolutely healthy_M
 ‘Vanya was absolutely healthy.’

2.4.2 Double gender nouns

In terms of agreement, double gender nouns allow a certain variability: both masculine and feminine agreement forms are available for the same lexical item, as shown in (8-a) and (8-b) respectively (from Crockett 1976:69).

- (8) a. Et-ot plaksa revel-Ø vsju noč.
 this_M crybaby howl_{PST.M} all night
 ‘This crybaby howled all night.’
 b. Et-a plaksa revel-a vsju noč.
 this_F crybaby howl_{PST.F} all night

Recall that in the case of double gender nouns the semantic feature has two values, i.e. ‘masculine’ and ‘feminine’ due to the fact that these nouns can refer to both males and females. Moreover, these nouns have feminine morphology (i.e. they belong to declension II where the majority of feminine nouns are found). This means that the mismatch only occurs when they

refer to males, otherwise the two features overlap. It is clear that in (8-a) masculine agreement has a semantic justification. In (8-b), on the other hand, feminine agreement can reflect either the semantic or the morphological property. In terms of gender assignment (Corbett 1991), double gender nouns take different agreements for semantic reasons: masculine when the referent of a noun is a male, as in (8-a), and feminine when the referent is a female, as in (8-b). Such motivation seems to be reasonable, since feminine agreement is not appropriate in the context of a male and masculine in the context of a female.¹¹

- (9) a. *Et-a plaksa revel-a vsju noč. (male referent)
 this_F crybaby howl_{PST.F} all night
 ‘This crybaby howled all night.’
 b. *Et-ot plaksa revel-Ø vsju noč. (female referent)
 this_M crybaby howl_{PST.M} all night
 ‘This crybaby howled all night.’

The data in (9) suggest that the variability observed for double gender nouns is obligatory in nature. These data also suggest that morphology might be redundant here, yet, syntactic (feminine) agreement is also attested for double gender nouns used in reference to a male. However, it is restricted to the construction with a copula verb, where an adjective modifying a double gender noun in the predicative (post-copula) position can show syntactic (feminine) agreement, while the pre-copula phrase is masculine. For example, in (10) the double gender noun *lakomka* ‘gourmand’ has a male referent, as indicated by the anaphoric pronoun *on* ‘he’ in the pre-copula phrase. Yet, feminine agreement in the post-copula phrase reveals that the adjective

¹¹As the question marks in (i-a) indicate, there are native speakers who hesitate to rule out the possibility of a male referent when agreement on the targets is feminine, yet this example shows very poor acceptability of a male. Furthermore, the possibility of a male referent is entirely ruled out for the structure in (i-b) where verb is the only target. The reason for this must be that adjectival modifiers are more prone to show variability than the verb. This will be evident in copular constructions with double gender nouns as well as with hybrids. Therefore, in the experimentation I decided to elicit verbal predicate agreement for double gender nouns, hybrids, and female names in *-ok/-ik*, where the likelihood of semantic agreement is very high.

- (i) a. ??Et-a plaksa revel-a vsju noč. (male referent)
 this_F crybaby howl_{PST.F} all night
 ‘This crybaby howled all night.’
 b. *Opjat’ plaksa revela vsju noč. (male referent)
 again crybaby howl_{PST.F} all night
 ‘Again the crybaby howled the whole night’

reflects the morphological feature of the double gender noun, but not the male sex of its referent.¹²

- (10) On - **izvestn-aja** lakomka.
 he well-known_F gourmand
 ‘He is a well-known gourmand.’

Semantic masculine agreement can also be used in the post-copula phrase, as shown in (11).

- (11) On - **izvestn-yj** lakomka.
 he well-known_M gourmand

In my experimental study I will focus on double gender nouns as subjects, similar to those in (8). Copular constructions, where double gender nouns are used as modifiers (hence they are not used referentially) are thus beyond the scope of this dissertation.

To conclude, despite the optionality in constructions with the copula, double gender nouns take two consistent agreement patterns that have semantic/referential justification, as shown in (12). This means that feminine markers are found on all agreement targets when the referent is a female, and masculine markers are found on all agreement targets when the referent is a male. Moreover this distribution holds in all case forms. These are the distinct characteristics of double gender nouns as compared to hybrids and female names in *-ok/-ik*, which are considered in the following sections.

¹²Crockett (1976:83) gives an example where masculine agreement is used with a double gender noun *zavodila* ‘livewire’ that has a female referent in the pre-copula phrase:

- (i) Umerl-a Nina Vladimirovna Smirnova - **naš-Ø škol'n-yj** zavodila.
 die_{PST.F} Nina Vladimirovna Smirnova our_M school_M livewire
 ‘Nina Vladimirovna Smirnova, our school’s livewire, is dead.’

Note, that in (i) the style of the sentence is rather formal and the noun *zavodila* ‘livewire’ is not used here to denote the behavior or personal qualities of the referent (as double gender nouns usually do), but rather the area of her school activity, somewhat like initiator or organizer, which brings this noun closer to the category of hybrids that can take masculine agreement for female referents (see Section 2.4.3).

Consider now the structure in (ii) which contains a type of a double gender noun which defines a female individual by her personal qualities:

- (ii) *Ona - **izvestn-yj** lakomka.
 she well-known_M gourmand
 ‘She is a well-known gourmand.’

- (12) a. Naš-emu plaks-e opjat' ne spalos'. (male)
 our_{M.DAT} crybaby_{DAT} again not sleep_{PST.refl.N}
 'Our crybaby again could not sleep.'
 b. Naš-ej plaks-e opjat' ne spalos'. (female)
 our_{F.DAT} crybaby_{DAT} again not sleep_{PST.refl.N}

2.4.3 Hybrids

Unlike double gender nouns, agreement manifestations with hybrids reveal variability which is optional in nature. Recall that hybrids is a class of declension I nouns that denote particular professions, e.g. *vrač* 'physician'. Like double gender nouns they can be used in reference to males or females in the same form, i.e. being masculine morphologically a hybrid does not have a feminine counterpart (*inžener* - **inženerka* 'engineer') and it is used with reference to a female in its masculine form. This creates a morpho-semantic mismatch.¹³

The social, cultural and political changes related to the status of women in society greatly affected the grammatical status of nouns denoting professional titles. Since the end of the XIX century females in Russia began to be employed in what were traditionally male dominated professions and received equal rights after the October Revolution of 1917. This caused significant changes in the gender system of the language. As a result, hybrids became sensitive to the sex of their female referents. This change is directly reflected in agreement.

First of all, syntactic (masculine) agreement is rather common in reference to both males and females, as shown in (13). In fact, when the referent is contextually unknown, the structure in (13) is ambiguous, since it can imply that either a male or a female physician came.

- (13) *vrač* *prišel* (a male or a female)
 physician come_{PST.M}
 'the physician came'

Secondly, semantic (feminine) agreement is also attested for a female referent, as shown in (14). According to Corbett (1991), it has become more frequent in the recent past. Clearly, semantic (feminine) agreement helps to avoid the ambiguity reported in (13).

¹³A few hybrid nouns have feminine counterparts, e.g. *vračixa* 'female physician' from *vrač* 'male physician'. However these forms have pejorative connotation and are not widely used in present-day Russian.

Hierarchy and, most importantly, the likelihood of semantic agreement is especially high for the verbal predicates as well as for relative and personal pronouns.

Finally, note that semantic (feminine) agreement is restricted to the nominative, therefore syntactic (masculine) agreement is found in the oblique cases in reference to a female:

- (20) Ja zapisalas' k *nov-oj / nov-omu vrač-u.
 I sign_{PST.F} up to new_{DAT.F} new_{DAT.M} physician_{DAT}
 'I made an appointment with a new physician.'

2.4.4 Female names in *-ok/-ik*

Unlike hybrids and double gender nouns female names in *-ok/-ik* denote only females, yet both semantic and morphological criteria can have an effect on agreement, which reveals optionality similar to that of hybrids. As the examples in (21) demonstrate, both syntactic (masculine) and semantic (feminine) agreement is attested with these nouns. In (21-a) masculine agreement expresses the morphological feature of the noun (to be more specific, the feature of the suffix *-ik*), while in (21-b) feminine agreement reflects the female sex of the referent.

- (21) a. Pomniš, Svetik byl-Ø tak-oj malen'k-ij.
 remember Svetik be_{PST.M} such_M little_M
 'Remember, Svetik was so little!'
 b. Pomniš, Svetik byl-a tak-aja malen'k-aja.
 remember Svetik be_{PST.F} such_F little_F

Similarly to hybrids, the variation in agreement involves only nominative singular, so that feminine agreement is disallowed in oblique cases (from Crockett 1976:53):

- (22) Pridodi so svo-im / *svo-ej Ninčik-om.
 come with your_{INS.M} your_{INS.F} Ninčik_{INS}
 'Come with your Ninčik.'

Importantly, these female names behave like hybrids referring to females, i.e. the choice of agreement appears to be constrained by the type of the target. For example, in (23) the feminine (semantic) form is found in the verbal predicate *byval-a*, however, the attributive adjective *ljubim-yj* (used post-nominally) is masculine, which matches the form of the noun *Svetik* (from Iomdin 1990:128).

- (23) A vot direktor ni razu ne podumal, čto Svetik ix **ljubim-yj**
 but director not once not thought that Svetik their beloved_M
 dal'še Irkutska ne **byval-a**.
 farther Irkutsk not be_{PST.F}
 'But the director never thought that their beloved Svetik had not
 been farther than Irkutsk.'

2.4.5 Summary

In the previous sections I have shown that both semantic and morphological criteria exert influence on agreement with the classes of nouns considered in this dissertation. In the majority of cases agreement has a semantic justification. Semantic agreement is obligatory for *papa*-type nouns, male names in *-a* and double gender nouns, which exhibit consistent agreement patterns (except copular construction for double gender nouns). In other words, the sex specification of their referents is obligatory reflected in agreement in all the targets. On the other hand, in the case of hybrids and female names in *-ok/-ik* semantic agreement is optional. It is constrained by the type of the agreement target. Semantic forms being most frequent in verbal predicates and relative and personal pronouns. Furthermore, semantic agreement is restricted to nominative singular only. Finally, agreement variability with these nouns is governed by socio-cultural factors.

The important question is how this complexity is manageable for native speakers. More specifically, how do first language learners know which principles are relevant for gender assignment in their language and which of them is dominant in which context? In the next section I briefly discuss two main approaches that deal with the problem of gender conflict resolution, such as assignment rules hierarchy (or semantic hierarchy) (Corbett 1991, Corbett and Fraser 2000) and the default (or markedness) hierarchy (Rice 2006). I show that neither theory can provide a straightforward explanation of the Russian data, nor do they pursue the issue of learnability in order to provide an explanation of how learning is actually achieved.

2.5 Semantic hierarchy vs. default hierarchy

From the start of his intensive investigation Corbett has been concerned with the native speaker's ability to allot nouns to gender (Corbett 1982; 1991, Corbett and Fraser 2000 *inter alia*). In his monumental work, Corbett (1991) formulates two sets of assignment rules necessary to establish a noun's gender: semantic, as in (24), and morphological, as in (25), which may form

part of the native speaker's competence.

- (24) Semantic assignment rules in Russian:
- a. Sex-differentiable nouns denoting males (humans and higher animals) are masculine.
 - b. Sex-differentiable nouns denoting females are feminine.
- (25) Morphological assignment rules in Russian:
- a. nouns of declensional class I are masculine;
 - b. nouns of declensional class II and III are feminine;
 - c. nouns of declensional class IV are neuter.

Furthermore, in order to account for the gender of nouns like *papa* 'daddy' Corbett postulates a hierarchy of gender assignment rules where the semantic rules take precedence over morphological rules.

- (26) Gender Hierarchy according to Corbett (1991):
 Semantic rules >> Morphological rules (>> = "outrank")

Clearly, nouns like *papa* demonstrate that the semantic assignment rules are dominant: masculine gender is assigned to these nouns by virtue of their semantics, more specifically, by the semantic rule in (24-a), which outranks the morphological rule in (25-b). While it is correct that in the case of *papa*-type nouns and male names in *-a* the conflict is resolved in favor of the semantically based principles, the argument may be weakened by other facts, which are considered below.

In a more recent work, Corbett and Fraser (2000:321) claim that semantics is the core of any gender system:

"As is universally the case, the formal assignment rules [...] are dominated by the semantic gender assignment rules."

Corbett and Fraser's claim about the universally dominant role of the semantic principles is based on typological evidence from over 200 languages, namely, there are languages where semantic information alone is sufficient for gender assignment (e.g. Godoberi, Tamil, among others), but there are no languages where formal information (phonological and/or morphological) is sufficient on its own. According to Corbett and Fraser, Russian is an example of a morphological assignment system "where semantic information is supplemented by morphological [...] information" (Corbett and Fraser 2000:294).

This idea seems rather paradoxical given that formal information allows correct gender assignment for the majority of Russian nouns. Thus it has

much larger scope than the semantic criterion. Moreover, the decisive role of the semantic rules is unambiguously evident only for the nouns like *papa* ‘daddy’ and male names in *-a*, which are rather few compared to the rest of the nouns.

The universal dominance of the semantic rules stated in (26) has been questioned by various researchers, who suggest that it might be too strong to refer to the general notion of the semantic rules as proposed by Corbett and Fraser (Dahl 2000, Nessel 2006). According to them the problem arises when one considers languages, like e.g. German, where formal rules cover a small proportion of the vocabulary. In these languages the researchers have postulated numerous semantic rules, some of which do not refer to biological sex and may be overridden by a morphological or phonological rule. For example, German superordinate nouns like *die Waffe* ‘weapon’ or *die Frucht* ‘fruit’ are feminine by virtue of the morphological rule (27-b) and the phonological (27-c) respectively (from Nessel 2006:1386, after Steinmetz 1986:190). This is problematic from Corbett and Fraser’s position, since the semantic rule in (27-a) loses in both cases.

- (27) a. Superordinate nouns are neuter (e.g. *das Möbel* ‘furniture’).
 b. Nouns in *-e* are feminine (e.g. *die Hose* trousers’).
 c. Nouns in /uxt/ are feminine (e.g. *die Bucht* ‘bay’).

Based on such evidence, Nessel (2006) has advanced the “Core Semantic Override Principle”, according to which only semantic rules invoking biological sex take precedence (from Nessel 2006:1386):¹⁴

- (28) The Core Semantic Override Principle:
 Rules referring to biological sex take precedence in gender assignment.

Yet, Russian data discussed in Sections 2.4.2, 2.4.3, and 2.4.4 seem to be problematic for the theories that postulate dominance of semantic or sex-based principles. In all three structures in (29) agreement has a morphological justification, which means that semantic/sex-based rules fail to assign gender to these nouns. It also means that morphological rules take precedence over the semantic/sex-based rules.

- (29) a. *vrač prišel* (female) (female)
 physician come_{PST.M}
 ‘the physician came’

¹⁴Previously Dahl (2000:102) pointed out that sex is the “major” gender criterion in languages with more than one gender for animates.

- b. Pomniš, Svetik byl-Ø tak-oj malen'k-ij. (female)
 remember Svetik be_{PST.M} such_M little_M
 'Remember, Svetik was so little!'
- c. On - izvestn-aja lakomka.
 he well-known_F gourmand
 'He is a well-known gourmand.'

The evidence in (29) suggests that it may be premature to draw conclusions on the question whether semantic/sex-based rules always take precedence in gender assignment.¹⁵ Some researchers (e.g. Steinmetz 1985; 1986, Rice 2006) argue that gender assignment constraints are equally ranked, so that semantics and morphology (or phonology) contribute to the process to the same extent. Nevertheless, as I show below, this view is also unlikely to account for all the facts.

Within the optimal gender assignment theory, Rice (2006) proposes that gender conflicts are resolved on the basis of the default or markedness hierarchy, formulated in (30) for Russian (from Steinmetz 1986:26):

- (30) Default Hierarchy for Modern Russian:
 Masculine >> Feminine >> Neuter (>> = "outrank")

The default hierarchy suggests that masculine is the default gender and neuter is the most marked gender. It can also be attributed to other languages that make a three-way gender distinction such as e.g. German. Consider now how the gender conflict for German superordinates like *die Waffe* 'weapon' is resolved in terms of the optimal gender assignment theory. Specifically, Rice (2006:1398) suggests that *die Waffe*, which falls into the domain of two constraints, one which assigns neuter to superordinates and one which assigns feminine to words in *-e*, is assigned to the least marked of the two conflicting categories, i.e. feminine.

With regard to the Russian data in (29), Rice's analysis makes correct predictions about (29-a) and (29-b), where the nouns *vrač* 'physician' and *Svetik* should be assigned masculine (in accordance with their morphology) and not feminine (in accordance with their semantics). This is consistent with the markedness hierarchy, where masculine is the least marked of the two conflicting categories. On the other hand, in (29-c) feminine agreement for the noun *lakomka* 'gourmand' suggests that the word should be assigned

¹⁵It should be noted that Corbett (1991) is aware of the problem with hybrids illustrated in (29-a). In fact it was he who proposed this term to capture the phenomenon of nouns like *vrač*, which he describes as a "... curious composite with one half being masculine, apparently half of a double-gender noun (when a male is denoted), while the other half is a hybrid noun (when a female is denoted)" (Corbett 1991:184).

feminine, which, however, is more marked than masculine. Thus, in some cases the default hierarchy is able to provide a better explanation of the intricate assignment patterns than the semantic hierarchy. Nevertheless, it appears that the gender system of Russian cannot be straightforwardly explained within one assignment theory, since the data discussed in this chapter do not seem to be fully compatible with the theories reviewed in this section. While these theories can deal with lexical gender of the nouns like *papa* ‘daddy’, referential gender of the nouns in (29) remains problematic for them.

Being concerned with finding the universally true theoretical approach to gender assignment, these theories do not pursue the issue of learnability.¹⁶ Yet, the question I would like to address is how the discovery of gender may proceed in the mind of a language learner. Another type of question that arises here has in fact been formulated by (Corbett 1991:82). “How does the child acquire the knowledge equivalent to that which we have modelled as assignment systems?” Further questions appear in the context of both the default or markedness hierarchy and the assignment rule hierarchy.

In terms of the default hierarchy approach, one might ask whether the unmarked form should be the easier and hence the earlier to be acquired. With the default hierarchy for Russian in mind (see (30) above), do we expect masculine gender to be acquired before feminine? And should neuter be acquired last? To put it differently, should children’s performance be superior for masculine nouns as compared to feminine and neuter? These are interesting research questions and to the best of my knowledge they have not received much attention in the research on child language in the context of gender in particular. However, exploring them further is not the focus of this dissertation and is therefore left as a topic for future research. Given that my main concern here is the discovery of the semantic principles in the course of acquisition and their integration into the system of grammatical gender, I would still like to mention here two interesting facts from Russian child language in relation to markedness. In the next chapter, where a detailed diary of one Russian child is reviewed, there is an indication that he initially develops a two-way gender system of masculine and feminine, while the acquisition of neuter gender is delayed. In addition, forms marking feminine gender are the first to appear in the child’s speech. However, these facts seem to be rather related to the properties of the input, such as transparency, consistency, and salience, and not a universal markedness hierarchy.

¹⁶(Corbett 1991:82) mentions child language acquisition as an area for further research, which, according to him, may provide a clearer picture of how assignment systems work.

Offering a detailed comparison of Corbett's and Rice's approaches is well beyond the scope of this dissertation. However, the questions arising from Corbett's assignment rule hierarchy are more relevant for my research. First, do children establish gender based on formal or semantic criteria? And what is the role of agreement in this process? More importantly, given that semantic (sex-based) criteria are on top of the assignment hierarchy, will they be hypothesized by children first? In fact such a scenario seems to be logically correct: children may want to initially rely on semantics, especially with regard to nouns like *papa* 'daddy', since otherwise they are doomed to commit errors, which will have to be unlearned. In fact, this idea, known under the term "semantic bootstrapping", was hypothesized by Pinker (1982) in one of his earliest works. He proposed that the discovery of gender may proceed as follows: the child may use "... the sex of human referents as a semantic cue for the feature name GENDER and the feature values *masc* and *fem*, with the gender of inanimate nouns learned distributionally via their similarity in inflection to words denoting humans". Yet, as I show in the next chapter, against all expectations previous research on child language reveals the opposite scenario, namely that children acquiring various languages base their initial hypotheses about gender on formal rules rather than semantic. Many issues related to acquisition of the semantic criterion and its role in the acquisition process remain unclear and will therefore be given special attention in this dissertation.

2.6 Summary of the chapter

This chapter has taken a detailed look at the main criteria of the gender system in Russian. In particular, with regard to the classes of nouns under investigation it has been shown that semantic and morphological criteria do not necessarily overlap, but often make conflicting gender predictions. Both of them exert influence on agreement, resulting in highly complex patterns. On the whole, the Russian gender system is transparent and consistent, but for the nouns considered in this dissertation gender assignment is complex and has inspired different theoretical approaches, as we have seen in Section 2.5. While a thorough discussion of these theories' relative merits are beyond the scope of this dissertation, the complexity of the system and the theoretical disagreement illustrate the need to find out more about how children acquire the problematic nouns. Once again, these nouns possess a number of important characteristics: (1) they have transparent but misleading morphological form; (2) sex-distinction is not necessarily an intrinsic property of the noun. For hybrids and double gender nouns it is established via identification

with a human referent in a particular discourse situation; (3) while semantic agreement is a constant grammatical property for some of these nouns (i.e. *papa*-type, male names in *-a*, double gender nouns (except the copula construction) it is only optional with the others (i.e. hybrids and female names in *-ok/-ik*).

The following chapter examines the main aspects of gender in relation to acquisition. Special attention is given to the findings from the acquisition of gender in conflict situations.

Chapter 3

Previous studies on the acquisition of grammatical gender

3.1 Introduction

Most of the previous research on gender acquisition by monolingual first language learners was carried out in the seventies and eighties and focused mainly on Romance, Germanic, Slavic, and Semitic languages (Karmiloff-Smith 1979, Clark 1985, Maratsos 1982, Popova 1973, Smoczyńska 1985, Henzl 1975, Mills 1986, Levy 1983a;b, Berman 1985 *inter alia*). The major concern of the researchers at that time were the following two big questions: (1) Do children begin to establish gender based on semantic (sex-based) criteria or are formal factors discovered before semantic? and (2) Is it possible for a child to acquire gender in a principled way? In other words, is gender acquisition a rote- or rule-based process?

The first question followed from an observation that the two gender criteria are different in nature. With regard to the semantic factor, there is a consistent correlation with meaning which involves a universally transparent notion of natural gender. The formal factor, on the other hand, is semantically abstract, language specific, and involves extremely diverse and complex patterns of inflection. In addition, the two factors come from two different sources: extra-linguistic and grammar-internal. In the eighties, the researchers who aimed to determine the order of acquisition of these gender principles were inspired by the idea of semantic bootstrapping:

“Initially, children do not have access to language form, but do have access to extra-linguistic forms of meaning. On the basis

of these meanings, children “bootstrap” to formal knowledge of language, i.e. to its forms and its units.” (Lust 2006:43)

In terms of gender acquisition it has been proposed that semantic criterion, which are part of the child’s mental structure, trigger the development of grammatical systems, so that the notion of natural gender, as an operating principle, must be applied to acquisition of formal gender distinctions (cf. Slobin 1973; 1985, Pinker 1982; 1984). However, this idea did not find empirical support. Instead it has been shown cross-linguistically that gender acquisition proceeds from formal-to-non-formal (cf. Karmiloff-Smith 1979, Henzl 1975, Levy 1983b). This point will be clarified on the basis of studies of French, Hebrew, Czech, and Russian considered in the following sections. Some researchers also tried to show that both gender criteria are acquired simultaneously (e.g. Mills 1986, Müller 2000). However, it is not completely clear when and how the semantic criteria are integrated into the gender system in the course of acquisition. More specifically, the question of how and when children realize that natural gender distinctions also play a role in the grammatical system of a language is still open.

The second question mentioned in the beginning raises the problem similar to what Corbett (1991) has been concerned with in his typological study, namely, whether native speakers remember the gender of every lexical item individually or whether they are able to predict it on the basis of gender relevant information. In the acquisition literature of the previous century and more recently it has been shown that children formulate language specific rules on the basis of formal information and apply them when determining and assigning gender, rather than learning the gender of every noun separately (more recently Müller 2000, Kupisch 2002). The order of appearance of some (formal) rules before others has been attributed to their transparency and consistency (cf. Mills (1986) for German, Gvozdev (1961) for Russian). Finally, the presence of (reliable) formal gender regularities has been shown to facilitate gender acquisition in monolingual and bilingual children (Kempe 2003, Kupisch 2002).

With regard to the semantic criterion, it has not been shown straightforwardly how children discover the knowledge of natural gender, and more importantly, how they integrate this extra-linguistic information into the grammar. Some researchers have observed that it is acquired gradually (and rather slowly) (e.g. Karmiloff-Smith 1979 for French, Gvozdev 1961 for Russian), and to my knowledge there is only one empirical study which focuses on gender conflicts in particular (i.e. Karmiloff-Smith 1979 on French). I would suggest that a gender conflict is the most fruitful domain for gaining deeper insights into the role of the semantic principle. I also suggest that

this linguistic domain can receive a novel, up-to-date interpretation when considered in light of two theories of morphological acquisition: the Words and Rules model (Pinker 1999) and the Rules and Competition model (Yang 2002), which consider the acquisition of regular and irregular (exceptional) morphology as dual vs. single process.

This chapter has the following organization. First, in Section 3.2 I consider some previous studies on the acquisition of gender in languages where both formal and semantic criteria are involved, notably French, German, Hebrew, and Czech. Previous findings from the acquisition of gender in Russian are discussed in Section 3.3, where I refer to both longitudinal (Gvozdev 1961) and experimental investigations (Popova 1973, Kempe 2003). Section 3.4 considers the findings from previous research on the acquisition of variable input, which may have implications for gender acquisition of nouns in variable vs. obligatory contexts. The summary of the chapter is presented in Section 3.5.

3.2 Semantic and formal aspects of the gender system through the prism of first language acquisition

Despite its complexity, the category of gender is in general acquired relatively early, by the age of 3 in many languages. Crosslinguistically, the time of acquisition differs and seems to be dependent on the transparency and consistency of the input or, in terms of gender assignment, on the straightforwardness of the system organization. For example, although Polish and Russian gender systems are rather similar, Russian gender is acquired somewhat later than Polish.¹ More specifically, according to Smoczyńska (1985), Polish children acquire a three-way gender distinction before the age of 2, while according to Gvozdev (1961), a Russian child does so by approximately the age of 3 (Gvozdev also finds that the gender of some Russian nouns is problematic even in late preschool years). Such a delay can be attributed to the fact that a Russian child faces a number of difficulties that a Polish child does not. One of them is the lack of correspondence between semantics and morphology. In addition, as mentioned in Chapter 2, the nominative singular form of some Russian nouns is opaque, e.g. those that end in a

¹Singular nouns in Polish display three genders: masculine, feminine, and neuter. In the plural, however, there is a two-way gender distinction: masculine personal and other (cf. Corbett 1991). The former applies to male humans, and the latter to all remaining nouns.

palatalized consonant (*sol'*_(F) ‘salt’ vs. *den'*_(M) ‘day’), or an unstressed *-o* (e.g. *zérkalo*_(N) ‘mirror’). Yet, as the acquisition data discussed in Section 3.3 demonstrate, this form is used as the main gender indicator by a Russian child between the ages of 2 and 3. The absence of such problematic cases in the input to a Polish child may explain why gender acquisition is precocious in Polish children.

With regard to other morphological gender systems, Maratsos (1982:249), for example, finds evidence that “German-speaking children formulate the noun gender system with surprisingly little difficulty” by the age of 3. Similar findings are discussed by Mills (1986) for German and by Levy (1983b) for Hebrew.

Contrary to what one might expect, semantic, i.e. natural gender systems, are not early acquired. Mills (1986), for example, reports that English three- and four-year-olds perform worse than German children of the same age with regard to gender-marked personal pronouns *he/she*. Corbett (1991:82-83) makes a similar informal observation about one English boy, who at the age of 4;2 did not use *he* and *she* in a target-consistent way. Specifically, “[h]e could distinguish males and females without difficulty, and used *he* consistently for males but sometimes also for females as well as *she*” (Corbett 1991:83). More interestingly, in De Houwer’s (1990) study, bilingual Dutch-English children acquired semantically determined Dutch pronouns before the English ones. Note that Dutch, unlike English, has a morphological assignment system. This means that not only is early gender acquisition possible where the system involves complex formal as well as semantic rules, such systems can be acquired before some straightforward semantic gender systems like e.g. English.² It should also be noted that the findings discussed above seem somewhat paradoxical from the standpoint of the semantic-bootstrapping hypothesis as well as from Corbett’s gender assignment model, where semantics is the core of the system. This important issue will be elaborated on in Chapter 9.

In the following sections I discuss the positions maintained with regard to child acquisition of gender based on the findings from French, German, Hebrew, and Czech. Special attention is paid to how children make use of different sources of information about gender in different languages, as well as to their ability to formulate rules.

²In fact there are indications in the literature that inflectional morphology (when it is transparent) can facilitate acquisition of various language systems: e.g. gender in Italian as compared to French (Kupisch 2002), determiners in Spanish as compared to Dutch (Baauw 2001), finite verbs in Italian and Brazilian Portuguese against English, German, and French (Kupisch to appear), etc.

3.2.1 French

As a part of her large study of determiners, Karmiloff-Smith (1979) carried out five experiments aiming to investigate the gender-indicating function of articles in French. The language has two genders, feminine and masculine, which are unambiguously marked on determiners in the singular (*le/un* for masculine and *la/une* for feminine). According to Kupisch (2002), French gender is predictable on the basis of the sound shape of nouns; however, formal rules are not sufficient for determining the gender of all nouns. Kupisch (2002:111) reports that “[m]ost word final consonants (71%) are associated with feminine gender ([k] constitutes an exception), while most word final vowels (75%) are associated with masculine gender ([i] constitutes an exception).” Nouns denoting males and females are generally masculine and feminine respectively. Note that form-gender mismatches occur in the language for nouns denoting humans as well as for inanimates. For example, nouns ending in a nasal vowel are usually masculine, yet exceptions like *femme* ‘woman’, *maison* ‘house’, etc. are feminine. Thus in some cases the phonological form of a noun is misleading.

Karmiloff-Smith (1979) conducted her experiments with 341 monolingual French children aged 3;2-11;11 to find out whether children use formal (phonological) or semantic criteria in determining the gender of novel nouns as expressed by the article (note that existing nouns were used in the introductory technique). Her main finding is that three-year-olds use the phonological information encoded on the noun’s suffix to determine the gender of the noun.³ She further claims that already at the age of 3 “the child constructed a very powerful, implicit system of phonological rules, based on the consistency, but not necessarily on frequency, of phonological changes in word endings” (Karmiloff-Smith 1979:167). Most interestingly for my study, she observes that neither syntactic information (i.e. the indefinite article provided by the experimenter) nor semantic information (i.e. the sex of persons depicted in the drawings) were determinant in eliciting gender agreement. Specifically, the lowest accuracy rates were obtained for the experimental condition where the elicitation phrase contained an indefinite article as gender indicator and a noun which carried an arbitrary suffix not providing a phonological clue, e.g. *un coumille* or *une dilare*. Thus, in the absence of semantic and phonological clues, the children, especially three- and four-year-olds, are not able to fully rely on the syntactic information provided by the article. In another experiment Karmiloff-Smith shows that children tend to derive gender information from the noun’s suffix and not from the

³Karmiloff-Smith uses the term *suffix* to refer to the word’s final segment. I keep her terminology in this section.

suffix of the indefinite article, since being provided with structures like *un goltine* and *une plichon* they erroneously produced *la goltine* and *le plichon*, where the forms of the definite articles match the form of the nouns. Correct answers would be *le goltine* and *la plichon*. In the cases of discord between the sex of the referent and the noun's phonological form, the children until the age of 10;0 predominantly rely on formal criteria in gender assignment. Yet, even the youngest children in her study are able to use semantic clues: three- and four-year-olds could use the semantic information in selecting the article when no phonological clue was available (i.e. the noun suffix was arbitrary) as well as in selecting personal pronouns. In addition, Karmiloff-Smith points out that children's consideration of both semantic and syntactic clues in conflict situations increases with age. Given these results, Karmiloff-Smith (1979:167) concludes that "... the phonological procedures are gradually (in some cases from 6 years, but more frequently at roughly 9 years) replaced by the natural gender clues and by the more foolproof syntactic ones ...".

The gradualness of the integration of the semantic and syntactic factors pointed out by Karmiloff-Smith is rather important. However, with regard to age it should be noted that the children she used are quite old and some of the tasks involving novel nouns may be rather hard for younger children. Therefore, there is a possibility that gender acquisition may proceed faster under natural conditions. In fact, according to Clark (1985) and Kupisch (2002), French children acquire gender by the age of 3 with little difficulty. Kupisch, for example, has found only five target-deviant structures with pronominal determiners in the speech of one French child aged 1;9-2;5. The errors contribute to 1.3% of all the structures she found. Among the errors which are often cited by the researchers are patterns with nouns that have a misleading phonological form, e.g. those that end in a nasal vowel typical of masculine nouns - *maman fait *un_M maison_(F)* 'mummy is making a house' (target structure: *maman fait une_F maison_(F)*) (from Clark 1985:706).

3.2.2 German

Mills (1986) has made an attempt to account for the acquisition of gender in two languages with profoundly distinct gender systems, namely English and German. While the former only makes a natural gender distinction reflected by the third person singular pronouns (*he, she, his, her*), the latter has a complex system including semantic and formal assignment principles with gender being marked on determiners and adjectives. The comparison of the two languages has been made in regard to the common feature, namely the male vs. female sex distinction of the referent. Surprisingly, Mills observes that German three- and four-year-olds are in advance of English children of

the same age in their production of gender-marked pronouns referring to sex-differentiated individuals. She attributes German children's advantage over English children to two factors. First, given that gender in German is marked on various agreement targets, children may be exposed to more evidence to acquire gender. Secondly, while in German pronouns *er* 'he' vs. *sie* 'she' are phonologically distinct, the similarity between the English *he*, *she*, and *her* may create problems for children. Mill's finding that gender in German is acquired somewhat easier than in English seems important. In Chapter 9 I suggest that on the crosslinguistic level this can be another indication that children proceed from formal-to-semantic in gender acquisition.

Another important observation made by Mills refers to the relationship of various rules, semantic and formal, within the gender system of a language. Based on observational data from three German children (aged 1;8-2;6) and elicitation data from German three- and four-year-olds, Mills proposes that rules are acquired in an order determined by the clarity of the rule in the particular language system. According to her idea, clarity of the rule is defined by the scope and number of exceptions: the greater the scope of the rule and the fewer the number of exceptions, the more quickly the rule will be acquired by the child. Most importantly, it is not the salience of formal rules over semantic or vice versa that constrains the order of acquisition but the clarity of a particular rule over another.

According to Mills, the first formal rule acquired by German children is the rule that assigns feminine gender to nouns ending in *-e*, since this rule "affects the largest part of the vocabulary, has the fewest exceptions and is clearly represented in the child's vocabulary" (Mills 1986:85). Other formal rules are acquired as the child's lexicon grows. Mills also suggests that "[t]he semantic rule of natural gender is also learnt early for certain forms", according to her by the age of 3 (Mills 1986:112). The evidence is found in both observational and experimental data. First, no errors occur in the speech of three- and four-year-olds who were asked to use personal pronouns in the subject function for the referent called by a proper name. This means that children always used *er* 'he' in reference to a male, and *sie* 'she' in reference to a female.⁴ In observational data she finds only two errors: **die_F Kind_(N)* 'the child' (target-structure *das_N Kind_(N)*) and **die_F Opa_(M)* 'the grandad' (target-structure *der_M Opa_(M)*). Both cases of erroneous production are very interesting and in fact may allow the researcher to make more specific claims about the course of gender acquisition than simply children's correct responses. The former indicates that the semantic rule was prioritized by the

⁴On page 103 Mills mentions that some children occasionally used both masculine and feminine pronouns in reference to the same person, which may indicate their indeterminacy.

child, since the referent of the noun *Kind* was a girl. (This may also indicate that the child is insensitive to the agreement evidence.) In the case of **die_F Opa_(M)*, both semantic and syntactic factors seem to be disregarded in favor of the phonological rule ‘die + N-[a]’ (formulated by Müller 2000:381). Interestingly, errors of this type are also observed by Müller (2000) in spontaneous speech of two-year-old German-French bilingual children. She reports that the children who have discovered the phonological rule ‘die + N-[a]’ can erroneously use it to assign feminine gender to the nouns that denote males, such as e.g. **die Opa* and **die Papa*. In reality, however, these nouns are masculine in accordance with the semantic rule, whereby nouns denoting males are masculine. At the same time children are found to produce the correct forms, i.e. *der Opa* and *der Papa*. Such overgeneralization errors suggest that studying children’s behavior in situations when formal and semantic criteria make conflicting predictions can shed more light on the relationship between formal and semantic gender principles. They also suggest that error patterns must receive special attention. These are central topics in the present study.

3.2.3 Hebrew

Similar to the findings from French and German discussed in the previous sections, two-year-old Hebrew-speaking children seem to put more weight on phonological principles than on semantic principles (Berman 1985, Levy 1983b). In particular, Levy (1983b), who analyzed her son’s (Arnon) spontaneous speech between the ages of 1;10 and 2;10 observes a gradual understanding of the phonological regularities of the final syllable of the noun. Note here that nouns in Hebrew are either masculine or feminine. Singular nouns ending in stressed /a/ or /t/ are typically feminine and take the suffix /ot/ in plural. All other nouns are generally masculine and take /im/ in plural. Agreement is expressed by adjectives, verbs in the past and future, as well as second and third person singular pronouns.

From an acquisition perspective, the gender system is rather straightforward, since, according to Levy, there is perfect correspondence between the sex of the referents of animate nouns and their grammatical gender. This means that nouns denoting males are masculine and nouns denoting females are feminine, and that in general, phonological and semantic criteria overlap. Levy suggests that evidence, such as children’s overgeneralization of natural gender distinctions to the inanimate objects that males and females typically own, may reveal their sensitivity to semantic regularities. However, no such evidence has been found. At the same time, some evidence revealing the non-semantic approach to gender acquisition has been found in the domain of second person pronouns *ata* ‘you’ (masculine, singular) and *at* ‘you’ (fem-

inine, singular). In terms of acquisition, the form *ata* used in reference to males can be problematic, as it is inflected with a typical feminine suffix /a/. Thus, if young children are sensitive to formal gender distinctions, they may erroneously use *ata* in reference to a female. Exactly this has been observed by Levy in her son's speech: Arnon, who is found to rely on phonological rule-based generalizations already at the age of 2;0, often fails to correctly address second person females. Levy demonstrates that the child performs at chance level, since only 202 (40.8%) out of 494 structures contained correct feminine pronoun *at* and a feminine verb form. This means that the child could erroneously use *ata* in reference to females as the form of the pronoun suggests feminine. Remarkably, Arnon shows perfect performance in addressing second person males with the 'feminine looking' pronoun *ata*. Levy attributes this finding mainly to the fact that the child received a lot more evidence for the use of *ata*, since both parents used it to address the boy and the mommy used it to address the daddy, while *at* appeared mostly in the speech of the daddy.

Erroneous choice of second person pronouns together with verb forms is taken by Levy as an indication that between the ages of 2 and 3 children lack "...cognitive clarity and cognitive salience of gender..." (Levy 1983b:91). In other words, the author suggests that children are unable to see that sex distinctions can have a grammatical realization. The author points out that errors due to the overgeneralization of formal features over semantic, which are repeatedly found in various child languages, present a special class of errors. Furthermore, as I show in this dissertation, their examination can be revealing of the nature of the processes underlying the acquisition of gender and other language systems.

3.2.4 Czech

Czech is a West Slavic language whose gender system is somewhat similar to Russian. Czech nouns are distributed between masculine, feminine and neuter genders. Noun gender can be predicted on the basis of inflectional morphology or on biological sex. Interestingly, there are cases where semantics and morphology make conflicting predictions, such as *děvče_(N)* 'a female child', *chlapisko_(N)* 'big man', or *děda_(M)* 'grandad'. The last example is identical to the Russian *deda_(M)* or *deduška_(M)* 'grandad' which has morphology typical of a feminine noun, but denotes a male.

The child data come from a series of elicited production experiments carried out by Henzl (1975) with three children aged 1;9-4;2. Both familiar and novel nouns were used in the study to elicit adjectival and verbal agreement. Henzl observes that children of all ages produced correct agreement for the

Czech nouns with transparent phonological form, i.e. nouns whose phonological form corresponds to their gender. For example, since *míč* ‘ball’ ends in a consonant, it is assigned masculine. However, in cases where the phonological form is ambiguous and does not correspond to the sex of the referent, errors occur. Importantly, errors with the noun *děda_(M)* ‘grandad’ are found at the ages of 1;9 and 2;10 for the children Olga and Dana. The same girls show correct masculine agreement with this noun when tested at the ages of 2;3 and 3;4 respectively. Based on these results, Henzl concludes that in the early stages, i.e. between the ages of 2 and 3, gender is assigned on the basis of the phonological form of the noun and not according to the noun’s semantics. Recall that similar observations have been made in the studies I have previously discussed. Moreover, Henzl suggests that children first formulate formal rules based on the noun suffixes which are least ambiguous rather than those which are most frequent. This seems to be in line with Karmiloff-Smith’s (1979) idea of consistency and Mill’s (1986) idea of clarity of the rule.

3.2.5 Summary

Based on the findings discussed in the previous sections the following main conclusions can be made:

1. Age 3 is the point that marks complete gender acquisition in many languages.
2. Morphological gender systems, even when they involve cases of form-meaning mismatches as well as other sorts of inconsistency and opacity, can be rather early acquired and in some cases earlier than semantic systems like e.g. English.
3. There are clear indications that two-year-olds possess an implicit system of formal rules. Little is known about the use of the semantic rule before the age of 3. Overgeneralization errors suggest that (a) children apply rules; (b) formal rules can interfere where the semantic rules are appropriate.
4. The order of acquisition of various formal rules seems to be affected by such factors as consistency and clarity, rather than frequency.
5. The semantic rule seems to be integrated gradually as children proceed from formal-to-semantic in gender acquisition.

6. Semantic criteria and especially formal ones are used by children as the main gender predictors.

In the next section I provide an overview of the main findings related to the acquisition of the Russian gender system in general and especially with regard to the noun classes that are the main target of this study. I also compare the relevant findings in Russian to those in other languages reviewed in this section and show that in general they are very similar.

3.3 Gender acquisition in Russian

Previous research on Russian gender consists of two empirical and one observational study. The latter is part of a detailed diary kept by Gvozdev (1961) of his son Zhenya from the child's first birthday, i.e. 1922, till the age of 9 years. Despite the remote date of creation, this study provides valuable insights into the development of the gender system. Popova's (1973) elicited production investigation was carried out in the 1960s and focused specifically on gender agreement with animate and inanimate Russian nouns. She tested 55 children aged 1;10-3;6 who attended two Moscow nurseries. Her results provide some interesting observations; however, many issues in her study are unclear (e.g. the way she presents the results by age and agreement type, as well as some methodological aspects, e.g. related to the question of when a certain structure is considered to be acquired). Therefore, her findings must be considered with caution. Finally, Kempe (2003) carried out an elicitation study with 46 monolingual Russian children aged 2;9-4;8 who attended a Moscow nursery. She specifically focused on children's agreement production with transparent simplex nouns ($dom_{(M)}$ 'house') and diminutive forms ($dom-ik_{DIM(M)}$ 'a small house') of familiar and novel nouns. Some of her findings that bear on the issues of familiarity and agreement type seem to be relevant in the context of my research. Thus, in what follows I mainly refer to Gvozdev's findings, at the same time implementing some relevant observations from Popova and Kempe. Throughout this section I draw parallels between Russian data and the data from other languages. Special attention is paid to the relationship of formal and semantic criteria in the course of acquisition.

3.3.1 Gender acquisition from onset till the age of 3

Gvozdev observes that between the age of 1;10 and 2, when Zhenya's sentences increase from two to three or four words in length, morphological

markers appear and grow fast on various grammatical categories such as number, case, and gender. As Gvozdev points out, the acquisition of gender lags behind the acquisition of number and case, which is attributed to the arbitrariness of the gender category and other factors. Interestingly, at some point Gvozdev expresses the idea that gender acquisition may be slowed down by the fact that children do not realize the grammatical value of sex distinctions at once. This is an important observation since natural gender bears on extra-linguistic knowledge which should be implemented in the grammatical system. Moreover, the number of cases where its independent role is obvious is rather small, while in the majority of cases formal and semantic features overlap, as shown in Chapter 2.

According to Gvozdev, the first nouns that appear in Zhenya's speech have the 'frozen' form of nominative singular, which is used predominantly also in the structures where other case forms are appropriate:⁵

- (1) Zhenya *pugovic-a *otorval-a (Zhenya 1;10.23)
 Zhenya_(M) button_{NOM} rip off_F
 'Zhenya ripped off the button.'
 Target structure: Zhenya_(M) pugovic-u_{ACC} otorval-Ø_M.

Shortly after 1;10 other case forms appear and start being used in a target-consistent way. Nominative singular is thus concluded to be the basic form, which strongly affects the acquisition of gender. Gvozdev observes that at around age 2;0, Zhenya develops a two-way gender distinction, i.e. masculine/feminine, while neuter is not present yet. This finding indicates that before the complete inflectional paradigm is acquired, the child tends to rely on the noun endings in the nominative singular and develops a nominative vs. feminine gender distinction based on the most transparent and most consistent noun endings, i.e. *-a* (typical for feminine nouns) and a hard consonant and zero ending (typical for masculine nouns). As mentioned in Chapter 2, nominative singular is not a reliable predictor for neuter nouns, since many of them end in an unstressed *-o*, which is identical to the unstressed *-a* due to vowel reduction. Obviously the child chooses the most reliable forms at the beginning, which supports the claim that the order of appearance of various formal rules is affected by the clarity and consistency of the noun suffix (cf. discussion of French, German and Czech above).

It should be noted that Zhenya goes through a very short stage (between the ages of 1;10 and 1;11) when only feminine agreement forms are produced

⁵The diary was recorded in a phonetic notation, so that the structure in (1) should be transliterated as *s'en' a buba ta' a*. The transliteration used here illustrates syntactic and morphological peculiarities only.

for all nouns irrespective of their gender and endings. In the following example we can see that feminine agreement on the verb is overgeneralized to a masculine noun that denotes a young male and has a transparent morphological form, i.e. it ends in a hard consonant:

- (2) mal'čik *letel-a tuda (Zhenya 1;10.19)
 boy_(M) fly_{PST.F} there
 'The boy flew there.'
 Target structure: Mal'čik_(M) poletel-Ø_{PST.M} tuda.

The evidence in (2) suggests that gender has not been discovered yet, rather the morpheme *-a* has been recognized and added on the verb without any sense of gender change. This phenomenon was called 'inflectional imperialism' by Slobin (1973). The choice of the morpheme *-a* in Russian is often explained in the literature (cf. e.g. Zakharova 1973) (1) by its acoustic salience, which makes it easy to be recognized by children, (2) by children's tendency to use open, prolonged syllables even with masculine nouns (e.g. children can say **tigr-a* instead of *tigr*_(M) 'tiger'), and (3) by the fact that around 70% of the child vocabulary at this stage consists of words ending in *-a*. A stage when feminine agreement is used exclusively is also reported by Popova (1973), but it is not clear how many children showed this behavior and during exactly what period (recall that the youngest children in her empirical study were aged 1;10).⁶

At the age of 1;11, according to Gvozdev, Zhenya enters the so-called mixed stage, when masculine and feminine agreement is used incorrectly even with nouns that have a transparent form which also corresponds to the sex of the referent:⁷

- (3) a. baba *pošel-Ø kuda (Zhenya 2;1.13)
 granny_(F) go_{PST.M} where
 'Where did granny go?'
 Target structure: Kuda pošl-a_{PST.F} baba_(F)?
 b. mal'čik *legl-a (Zhenya 2;0.10)
 boy_(M) lie_{PST.F}
 'The boy lay down.'
 Target structure: Mal'čik_(M) leg_{PST.M}.

⁶In general, the way the children are grouped in Popova's study and the fact that some of the groups overlap make it hard to interpret the results.

⁷Unlike Gvozdev, Popova distinguishes one more stage before the mixed stage, which she characterizes as 'predominance of masculine gender'; however, the evidence supporting this claim is rather unclear.

- (6) a. ja čital-Ø (male speaker)
 I read_{PST.M}
 ‘I was reading.’
 b. ja čital-a (female speaker)
 I read_{PST.F}
 ‘I was reading.’
- (7) a. ty čital-Ø (male addressee)
 you read_{PST.M}
 ‘You were reading.’
 b. ty čital-a (female addressee)
 you read_{PST.F}
 ‘You were reading.’

Clearly, it is the sex of the referent that is reflected through agreement in these cases; the forms of these pronouns do not carry any gender relevant morphological markers, and there is only one form for both males and females.

Gvozdev finds numerous errors with these pronouns between the age of 2 and 3 together with target-consistent forms.

- (8) a. *ty xotel-a (about father) (Zhenya 2;6.2)
 you want_{PST.F}
 ‘You wanted.’
 Target structure: Ty xotel-Ø_{PST.M}.
 b. *ja rasskazyval-a (about himself) (Zhenya 2;11.7)
 ja tell_{PST.F}
 ‘I was telling.’
 Target structure: Ja rasskazyval-Ø_{PST.M}.
 c. ty sam-Ø (Zhenya 2;5.9)
 you self_{PST.M}
 ‘You yourself.’

According to Gvozdev, agreement with *ty* ‘you’ is fully mastered by the age of 2;8, while for *ja* ‘I’ it is not acquired until the age of 3, when the author finds the last example of feminine agreement that the boy uses talking about himself. The fact that first and second person pronouns are problematic in terms of gender becomes obvious when one considers the evidence in the input to a Russian child, since the child hears caregivers using both masculine and feminine agreements with these pronominal forms.⁸ Before I continue

⁸For third person pronouns *on* ‘he’ and *ona* ‘she’, Gvozdev mentions that verbal agreement with them is acquired at the age of 2;1. However, when these forms (more frequently *on*) are used anaphorically, errors occur till approximately the age of 3. For example, Zhenya can erroneously use *on* ‘he’ for *sobaka*_(F) ‘dog’ (2;5.25) or *ona* ‘she’ for *karandaš*_(M)

the discussion, I would like to bring up an important issue concerning Gvozdev's study. Given that his findings were observed in the form of diary notes, it is not possible to know exactly how frequently certain structures occurred in Zhenya's speech at a particular age. Lacking this information means that it is not possible to draw reliable conclusions about whether some target-deviant forms were performance errors or part of the child's grammar. Therefore, the elicited production technique is used in this study to show which structures are part of the child's grammar and which are not.

Before Zhenya starts using first person pronoun *ja* he addresses himself as *mal'čik* 'boy' or by using his name *Zhenya* (a short form of his full name Evgenij), as shown in (1), (2), and (3-b) above. In all of them target-deviant agreement forms are found, which is consistent with the mixed developmental stage proposed by Gvozdev. In terms of meaning-form mismatches, the example in (1) is of more interest for me, since the form of this proper name, which suggests feminine, does not correspond to the sex of the person it denotes, i.e. a male. After the child starts using the pronoun *ja* there are no more instances of *Zhenya* being used about himself. Gvozdev does not mention any errors with other male names in *-a* in Zhenya's speech. It may be the case that they were assigned correct masculine gender in accordance with the semantic principle. However, it is also possible that Zhenya simply did not talk about other males by using their names. In other words, there may not have been any relevant examples.

By age 3 Zhenya hardly makes any mistakes with *papa*-type nouns. The target-deviant structure in (9) presents the last occurrence.

- (9) u *lenin-oj pap-y (Zhenya 2.9.21)
 at Lena_{GEN,F} daddy_{GEN(M)}
 'Lena's daddy has ...'
 Target structure: U Lenin-og_{GEN(M)} papy.

Similarly to Gvozdev, Popova observes occasional errors for this noun class in the speech of the children who show full mastery of gender assignment principles for masculine and feminine nouns with transparent morphological form (the exact age of these children is rather unclear). Interestingly, in one of the two instances that she reports, the form of the adjective reveals correct masculine agreement, while the form of the verb shows target-deviant feminine agreement (see (10-b)).

'pencil' (2;2.8).

- (10) a. *djadja* **sidel-a* na lošadke
 uncle_(M) sat_F on horse
 ‘The uncle sat on a horse.’
 Target structure: *Djadja*_(M) *sidel*_M na lošadke.
- b. *moj-Ø* *papa* **zabolel-a*
 my_M daddy_(M) get sick_{PST.F}
 ‘My daddy got sick.’
 Target structure: *Moj-Ø* *papa*_(M) *zabolel-Ø*_{PST.M}.

The type of agreement target appears to be an interesting and important issue in Gvozdev’s study as well. Specifically, he observes that verbal and adjectival forms do not emerge in Zhenya’s speech simultaneously. The first instances of noun-verb agreement occur at the age of 1;10 and noun-adjective agreement appears two months later, i.e. at the age of 2;0.⁹ Given this time gap, verbal agreement seems to be acquired somewhat earlier than the adjectival. In (11) there is a curious example where correct feminine agreement on the verb is followed by target-deviant masculine agreement on the adjective, which means that the child is more confident about subject-verb agreement than about noun-adjective agreement:¹⁰

- (11) *legl-a* (and next to it) *sobačka* **malen’k-ij* (Zhenya 2;1.12)
 lie_{PST.F} dog_{DIM.(F)} little_M
 ‘A little dog lay down.’
 Target structure: *Sobačka*_{DIM.(F)} *malen’k-aja*_F.

According to Gvozdev, the situation stabilizes by the age of 2;4. Even for first and second person pronouns the last target-deviant adjectival forms occur at the age of 2;5.

The most obvious explanation for the time difference between the acquisition of verbal and adjectival agreement, as proposed by some Russian researchers (e.g. Voeykova 1997), is the inflectional complexity of adjectival full forms (functioning both as predicate and attribute). That is, when full forms of adjectives are considered, their inflections are long, often disyllabic, and involve lexical allomorphy. For example, in nominative singular, masculine forms can take *-yj* or *-oj* and feminine *-aja* or *-ija* (cf. also Table 2.2 in Chapter 2). On the other hand, verbal inflections are shorter and have fairly simple morphology. The difference between masculine and feminine forms in

⁹The first adjectives are used predicatively and occur after the noun, but even when attributive forms appear they often occur in post-nominal position.

¹⁰Note that the example in (10-b) above shows the opposite, i.e. correct agreement on the possessive pronoun in attributive position and erroneous agreement on the verb. Unfortunately, the age of this child is not clear from Popova’s (1973) study.

nominative singular is rather straightforward: $-\emptyset$ vs. $-a$ respectively. Note also that there is no part of the adjectival paradigm that does not require a morpheme.

Kempe (2003), who compared the proportions of errors for various agreement types, finds that adjectival agreement is more error-prone than agreement between nouns and third person singular pronouns (*on* ‘he’ and *ona* ‘she’). Specifically, the children in her study (age 2;9-4;8) produced 10.4% (27/258) target-deviant structures for adjectival agreement and 5.5% (43/769) for pronominal agreement.¹¹ No errors occurred in children’s production with verbal agreement; however the number of verbal forms is very small (they were only 11) as compared to pronouns and adjectives. I suggest that although Kempe’s children perform somewhat worse with adjectives than with third person pronouns as well as verbs, no specific conclusions can be made based on these results, since the sample sizes differ greatly: a total of 11 items for verbal agreement vs. 258 for adjectival, and 769 for third person pronouns.

Another of Kempe’s findings that can be of interest for the present study is the noun familiarity effect. Specifically, based on the observation that gender agreement errors are more common for novel nouns than for familiar nouns, she claims that item-based learning is involved, i.e. certain agreement patterns are associated with individual nouns. At the beginning of this chapter I pointed out that the researchers concerned with formal aspects of gender acquisition reject the idea of item-based learning, since the evidence strongly suggests that at an early age (approximately the age of 2) children’s grammars have an implicit system of morpho-phonological gender regularities (cf. Mills 1986, Levy 1983b *inter alia*). Yet, this implication will be considered further in the context of natural gender acquisition. In particular, in Chapter 6, children’s agreement production with familiar, rare, and novel nouns is tested against two specific hypotheses formulated within the Words and Rules model (Pinker 1999) and the Rules and Competition model (Yang 2002). It will be shown that frequency of occurrence can have an effect on the course of acquisition with *papa*-type nouns, suggesting that a rote-based process is involved.

I now turn to some other problematic cases that Gvozdev observes between the ages of 2 and 3, namely neuter nouns, in particular those ending in unstressed $-o$ in nominative singular, and masculine and feminine nouns ending in a palatalized consonant in nominative singular. As mentioned in

¹¹It seems to me that the percentage calculations suggested by Kempe are not correctly performed. She reports 4.0% errors for personal pronouns and 2.5% for adjectives (cf. Kempe 2003:479).

The avoidance strategy, somewhat similar to the one in (13), is also applied to nouns ending in palatalized consonants. In (15) the child adjusts the sound-structure of the noun *step'* 'steppe' into *stepj-a*, so that it becomes phonologically identical to the feminine nouns ending in *-a*. This example demonstrates that the child prefers a phonologically clear but non-existent form.

- (15) malen'k-aja *step-ja (Zhenya 5;5.13)
small_F steppe
 'The small steppe.'
 Target structure: Malen'k-aja_F step'_(F).

For a long time, even after age 3, Zhenya erroneously assigns masculine gender to feminine nouns, apparently being confused by the palatalized endings of these nouns. (Interestingly, Gvozdev reports no errors of the opposite, i.e. masculine nouns like *den'_(M)* 'day' becoming feminine.)

- (16) ten' *upal-Ø (Zhenya 4;7.16)
shadow_(F) fall_{PST.M}
 'The shadow fell.'
 Target structure: Ten' upal-a_{PST.F}.

Based on the overgeneralization errors discussed in (12), (13), (15), and (16) two important conclusions can be made: (1) the child does not rely on the entire inflectional paradigm, but on the nominative singular case form, since if all the case forms were taken into consideration, he would realize the difference between declension I and IV, and II and III; (2) gender information is mainly derived from semantic and morphological criteria, but not from the agreement manifestations. The latter is especially obvious with regard to nouns ending in palatalized consonants. While the phonological form of masculine and feminine nouns ending in palatalized consonants is the same in the nominative singular, the sound pattern of the agreeing forms presents a straightforward contrast:

- (17) a. Kon' upal-Ø. vs. Ten' upal-/ə/.
horse_(M) fall_{PST.M} shadow_(F) fall_{PST.F}
 'The horse fell.' vs. 'The shadow fell.'
 b. Moj-Ø kon'. vs. Moj-á ten'.
my_M horse_(M) my_F shadow_(F)
 'My horse.' vs. 'My shadow.'

While agreement information appears to be a truly reliable clue for nouns ending in palatalized consonants, it does not always straightforwardly express

Both structures in (20) illustrate an attempt to match the form of the noun with the semantic properties of their male referents. This is successfully achieved by omitting the ending *-a*, typical of feminine nouns (from Čukovskij 1965:375).

- (20) a. bednyj zajčonok tebjja ***pjanic** sbil
 poor hare you_{GEN} knock down_{PST.M}
 ‘Poor hare, a drunkard knocked you down.’
 Target structure: Bednyj zajčonok, tebjja pjanica sbil.
- b. papa ***bjak** ‘Daddy is a baddie.’
 daddy_(M)
 Target structure: Papa - bjaka.

3.3.2 Summary of previous research on Russian

1. The course of acquisition seems to consist of at least three stages: (a) 1;10-1;11: predominance of feminine forms in agreement; gender distinctions are not discovered yet; (b) 1;11-2;4: mixed stage - masculine and feminine forms are used randomly; (c) 2;4-3;0 two-way (masculine/feminine) gender system is established. Full mastery of neuter nouns and nouns ending in palatalized consonants takes place after the age of 3.
2. Initially, children orient themselves to the clearest and most consistent formal patterns, e.g. that nouns ending in *-a* in the nominative singular are feminine, and nouns that have a zero ending are masculine.
3. Nominative singular is used as the basic form in determining the noun’s gender.
4. Although by the age of 3 first and second declension nouns (including those that denote human beings, but not the nouns ending in palatalized consonants) are correctly assigned masculine and feminine gender, it cannot be argued that the semantic principle is operative. Since in the majority of cases semantics and morphology overlap, it is not possible to say that the former was used, especially given that young children are very sensitive to formal gender regularities. With regard to form-meaning mismatches, the evidence is rather poor. The formal-to-semantic sequence seems evident for the noun *papa* ‘daddy’, yet it is

communication from parents and later published in a rather informal way.

only based on a few occurrences, and the exact accuracy rates are unknown. Finally, based on the agreement production for first and second person pronouns *ja* 'I' and *ty* 'you', it can be stated that consideration of semantic properties increases with age.

5. The gender of nouns is mainly derived from morphological criteria, rather than from semantic information or from agreement manifestations.

3.4 Previous research on children's acquisition of variation in the input

From the discussion of the previous research on gender acquisition it is clear that the focus of interest has been on how children acquire the grammatical structure of a language, in other words, on how they acquire the categorical forms. However, the acquisition of a gender system like Russian presents additional interest because the adult system allows for variability. As discussed in Chapter 2 Sections 2.4.3 and 2.4.4, variable forms are attested for two subcategories of nouns: hybrids referring to females and female names in *-ok/-ik*.¹⁴ This phenomenon, as I said, has been caused by socio-cultural factors, such as participation of women in social, cultural, and other spheres of life.

Variation in the input draws more and more attention in language acquisition research. Therefore, in this section I discuss some relevant findings of recent acquisition studies on variable input. It should be noted that despite great interest relatively little has been done on how children treat optionality in the input. Recently, the research on this subject has focused on the acquisition of phonological (e.g. Roberts 1997a;b; 2002, Smith 2007) and morphosyntactic variables (e.g. Smith 2007) in different languages. It has been demonstrated that children exposed to variable input are variable in their own production. However, the constraints that govern variability in the adult language are not acquired simultaneously. Most importantly, it has been argued by Roberts (1997a) and Smith (2007) among others that extra-linguistic (social) constraints can be acquired later than linguistic (grammatical) ones.

There is growing interest to language-internal variation in other areas of grammar, e.g. word order variation. Lately, some work has been done on the acquisition of object shift construction in various languages (Swedish

¹⁴Another variable pattern was attested for double gender nouns used predicatively in copular constructions, which are beyond the scope of my investigation.

(Josefsson 1996), Dutch (Barbier 2000, Schaeffer 2000)) where children's optional behavior is sometimes accounted by the lack of 'concept of non-shared knowledge' (Schaeffer 2000).¹⁵ Furthermore, the work in progress on the acquisition of object shift vs. subject shift in Norwegian child language has revealed some asymmetries for the two constructions which are accounted for in terms of structural economy (Westergaard 2007).

With regard to the present study, Smith's (2007) findings from the acquisition of certain morphosyntactic variables in a Scottish dialect seem to be especially interesting. Smith reports that children (2;10-3;6) acquiring a Scottish dialect (Buckie) respect the constraint on the use of verbal *-s* in third-person-plural contexts from the start of the acquisition process. According to Smith (2007:80), the marker *-s* on the verb is used variably in plural noun phrases (NPs) in Buckie, but never or very rarely with the pronoun *they*. Note that while the use of *-s* with plural NPs is grammatical in Buckie, it goes against the rules of Standard English. Hence the use of *-s* with plural NPs is most expected within the community, especially in informal situational contexts of play or daily routine vs. formal contexts, such as teaching or discipline, where the standard norm is more likely to occur.¹⁶ Based on a comparison of the child, community and caregiver production, Smith (2007) suggests that two- and three-year-olds have knowledge of categorical vs. variable rules, since they use *-s* variably with full NPs, but almost never with the pronoun *they*, similarly to adults. However, the children in her study fail to acquire the external stylistic constraints on its use. That is children do not show distinction in use of the local variant according to function: informal speech variant, i.e. '*-s* with plural NPs, in informal contexts and formal speech variant, i.e. no *-s*, in formal contexts.

The delay in the acquisition of stylistic constraints as compared to language internal (grammatical) constraints is ascribed to the nature of caregiver speech. Specifically, Smith suggests that since caregivers are not consciously aware of the social impact of the variants they are using, their children are not aware of them either. According to her, caregiver's awareness of the social weight of the variable forms has an impact on when they are acquired. She suggests that caregivers can "teach" their children the sociolinguistic rules by changing their linguistic behavior in adult-to-adult vs. caregiver-child interactions in order to show to the child when and where it is appropriate to use one form or the other. Since in the case of the verbal *-s* variable the

¹⁵Object Shift is a syntactic operation that raises unstressed pronominal objects past the negation or a sentence adverbial. A similar construction, called Subject Shift, presents a leftward movement of pronominal or light subjects past the adverb or negation. Both of these constructions are found in Scandinavian languages among others.

¹⁶These formality constraints have been proposed by (Labov 2001:437).

adults do not make a distinction between the standard norm and the dialect form explicit to children, it will take time for them to realize that one form is widespread outside the community and the other one is local.

Most importantly, the studies reviewed above converge in showing that sociolinguistic competence can be acquired later than grammatical competence. In other words, often despite their complexity, grammar-internal characteristics of a particular phenomenon can be acquired before its extra-linguistic properties. This finding can have an interesting implication for the acquisition of the Russian gender system, where in some contexts, which I refer to as obligatory contexts, the semantic rule is categorical (e.g. with *papa*-type nouns), while in other contexts - variable contexts - it is variable (e.g. with hybrids referring to females and female names in *-ok/-ik*). Most importantly, this variation is socially motivated. Thus, the former case represents grammatical knowledge and the latter sociolinguistic awareness. It is thus important to explore children's agreement behavior in obligatory vs. variable contexts.

3.5 Summary of the chapter

In this chapter, I have reviewed various studies on the acquisition of gender in Russian and other languages, focusing on the cases where formal and semantic gender principles are in conflict. Gender acquisition has been considered in relation to important questions, such as when and how the semantic principle is integrated into the gender system in the course of acquisition, which appears to be largely determined by the formal criteria at the beginning. Another important issue that has been discussed in this chapter is related to the acquisition of variable forms. One aspect of it, i.e. the order of acquisition of grammatical vs. sociolinguistic constraints on variation, received special attention.

In the next chapter, based on findings from previous studies outlined here as well as theoretical considerations discussed in Chapter 2, I formulate the research questions and predictions for this empirical study.

Chapter 4

General predictions and research questions

4.1 Introduction

Based on the crosslinguistic evidence and in particular the relevant findings from Russian gender acquisition discussed in Chapter 3 as well as on some important theoretical assumptions considered in Chapter 2, I now formulate the research questions with regard to the classes of nouns considered in this dissertation. In the previous chapter we have seen that the course of acquisition depends on clarity and consistency of the linguistic input to which a child has been exposed. We have also seen that exceptions or counterexamples to clear and consistent data make the acquisition task harder, as reflected in children's non-adult-like behavior. From an acquisition perspective, the nouns considered in this dissertation are exceptional, as they fall within the domain of two competing gender principles: morphological and semantic. Yet, little is known about how children behave with these nouns, especially with hybrids, female names, and double gender nouns. In other words, there is a dearth of information about how children establish the adult-like relationship between the morphological and semantic gender criteria.

From the acquisition data reviewed we have seen that children's initial hypotheses about gender in various languages are based on generalizations from the formal properties of a noun, which does not always result in the right outcome. In Russian, the morphological properties of the nouns under investigation appear to be overgeneralized between the ages of 2 and 3, as shown for the noun *papa* 'daddy' and some male names in *-a* in Gvozdev's (1961) diary. As a Russian child approaches the age of 3, morphological-rule-based

generalizations seem to be gradually replaced by those made on the basis of the semantic concept of natural gender. Some evidence showing that the integration of the semantic concept may be the result of development comes from the noun *papa* ‘daddy’ as well as from first and second person pronouns (cf. the discussion in Chapter 3 Section 3.3). (Recall also that Karmiloff-Smith (1979) noted a gradual replacement of a phonological procedure by the semantic one for the novel nouns denoting humans in French.) However, very little is known about children’s behavior with other noun classes. Therefore, they should be taken into consideration, especially given that not all nouns are acquired uniformly by Zhenya (cf. the acquisition of neuter nouns and nouns ending in palatalized consonants).

The main research question the present study aims to answer is how the knowledge of gender grows over time. Furthermore, a number of more specific questions can be formulated based on the findings summarized above as well as some other important theoretical and empirical issues. The first set of questions is related to the gradualism observed in the acquisition of the semantic principle in a Russian child as well as in the children acquiring other languages. In particular it is relevant to investigate exactly how the integration of the semantic principle proceeds. Is the change from formal-to-semantic transitional or abrupt? Are there any differences for individual nouns? If yes, what are they and what underlies the pattern? These questions bear on the notion of frequency which has been a heavily debated subject in language acquisition research in the past few years (Yang 2002, Tomasello 2003, Theakston et al 2004 *inter alia*). Hence, in Section 4.2 in the context of a more general discussion of frequency issues I review two theories of morphological acquisition - the Words and Rules model (Pinker 1999) and the Rules and Competition model (Yang 2002) - and show how their assumptions can be generalized to the acquisition of gender with *papa*-type nouns. This discussion lays the groundwork for the study of gender acquisition in obligatory contexts, i.e. for *papa*-type nouns and male names in *-a*, which is presented in Chapter 6.

Another set of questions is related to the acquisition of variable forms, which is relevant for the acquisition of hybrids and female names in *-ok/-ik*, where the semantic rule is used variably and not categorically (as with *papa*-type nouns and male names in *-a*). The questions to ask here are the following: How do children deal with variation in the input? Is the semantic rule used variably for female referents? Do children exhibit any preferences in such cases and does their behavior match that of the adults? What kinds of considerations influence their choices? Does the acquisition of the semantic principle take place simultaneously in variable vs. obligatory contexts? These questions will be addressed in Chapter 7 where I investigate children’s ability

to make use of the semantic rule with hybrids referring to females and female names in *-ok/-ik*. The predictions that I put to test there are based on the previous research on the acquisition of variable forms and more specifically on the acquisition of grammatical vs. socio-cultural constraints on variability discussed in Chapter 3 Section 3.4. These predictions are presented in Section 4.3.

With regard to the last subcategory of nouns, i.e. double gender nouns, which will be investigated in Chapter 8, I address the following questions: How is the semantic principle acquired with double gender nouns where the same noun form is used to refer to both males and females? Do children have knowledge of referential gender? In other words, do they know that the gender of a double gender noun is decided for every single noun form in a concrete discourse situation and that it must correspond to the biological sex of the specific referent? These questions give rise to several predictions regarding the referential properties of double gender nouns, which are discussed in Section 4.4 and elaborated upon at the beginning of Chapter 8.

The final set of questions, which aims at summarizing the results presented in this study, concerns the establishment of the semantic principle across different noun classes. The questions at issue are the following: When children recognize the role of semantics, do they do so for all nouns simultaneously? In other words, when children realize the grammatical function of natural gender do they apply the semantic rule across all noun classes to which it is relevant for? In Section 4.5 I will formulate two alternative predictions with regard to the acquisition of the semantic rule across the noun classes considered in this dissertation. These predictions will be kept in mind throughout the chapters of analysis of the individual subtypes of nouns. In Chapter 9 the results of the experimentation will be summarized and reconsidered regarding these predictions.

4.2 Frequency issues

Whether or not one assumes that there is an innate component called Universal Grammar (UG), primary linguistic data, or linguistic input, is essential for acquiring a language. Recently, the effect of input on the acquisition process has received considerable attention. Many researchers have felt the necessity to explore the role of input in the acquisition of various grammatical phenomena, due to asymmetries in child grammatical development observed on the crosslinguistic as well as language internal level (Kupisch 2007, Westergaard and Bentzen 2007, among others in the volume).

From a theoretical perspective, two opposing views on input exist: while

in various constructivist approaches, input plays a major role (Tomasello 2003, Theakston 2004), in the generative theory of Universal Grammar (UG) in its classical form (Chomsky 1965) the role of input is minimal. Within the constructivist framework it has been proposed that children form grammatical hypotheses based entirely on the input. Specifically, children are assumed to build abstract linguistic constructions based on the imitative learning, i.e. in an item-by-item fashion, gradually improving their learning skills in the course of development. From a generative perspective, however, input plays a limited role in grammar formation, as children formulate grammatical hypotheses based on innately specified linguistic knowledge. Input or, more specifically, pieces of input containing unambiguous cues, are said to “trigger” the acquisition process (Lightfoot 1999). Yet, the relevance of input frequency has been brought into focus in cases where UG provides parametric alternatives (e.g. subject omission (Pro Drop), head direction (head initial vs. head final languages), etc.). Recently, Yang (2002) has proposed a parameter setting approach, the Variational Model, according to which the grammars defined by UG compete against each other. According to Yang (2002), input frequency is a deciding factor which explains a gradual transition from one grammar to another. Importantly for us, Yang extends the variational learning model to the acquisition of morphology, in particular English past tense. His model, called the Rules and Competition model (henceforth the RC model) contrasts with another leading approach developed by Pinker (1999), which is known as the Words and Rules model (henceforth the WR model). While both models attribute the major role to frequency in morphological acquisition, they make different claims regarding the mechanisms that explain how learning is actually achieved. The WR model proposes a dual mechanism approach where regular morphology is acquired by rules, while irregular morphology is stored in the mental lexicon. The RC model proposes a single mechanism approach where both regulars and irregulars are acquired by rules. In the following subsections I provide the description of important features of the WR and RC models and extend these ideas to the acquisition of gender in the context of the exceptional noun classes.

4.2.1 The words and rules model

The WR model (Pinker 1999) is a dual mechanism theory where the two components ‘word’ and ‘rule’ represent two processes: rote learning for irregular morphology and rule learning for regular morphology. Hence in English regular past tense verb formation is subject to a rule application, namely the default phonological rule which adds the *-ed* suffix to the root (stem).

Irregular verb forms, on the other hand, are claimed to be stored in the mental lexicon. That is, since these forms are unpredictable, they must be memorized or learnt by rote. Importantly, Pinker proposes that irregulars are not stored at random, but rather according to patterns. This means that lexical items with similar patterns are stored close together, e.g. *sing-sang* is stored in the neighborhood of *ring-rang*¹. The WR model predicts that a child will initially memorize irregular forms of individual verbs, later she will discover patterns among these different items in adult usage and generalize them to new, similar verbs. Thus, the acquisition of irregular morphology starts as a rote-based process and later proceeds in a rule-based fashion. Token frequency is the crucial factor that affects this process. The model posits that irregular forms are memorized with a certain strength based on token frequency, i.e. the frequency of occurrence of an individual verb in the input. High token frequency of a verb enables it to be learnt faster and with greater accuracy. Since memorization takes time and experience to be perfected, the child can fail to retrieve an irregular form, and in this case the default *-ed* form will be used. According to Pinker, the more frequently an irregular verb is heard, the better the memory retrieval for that verb gets, and the lower the overregularization rate.

4.2.2 The rules and competition model

Unlike Pinker, Yang (2002) introduces a single component, i.e. a ‘rule’ component. In his model English irregular past tense forms, as well as regular ones, are formed by rules that apply to a class of individual verbs (e.g. vowel shortening *feed, shoot, ...*). Thus, in the RC model it is the class membership that is memorized. The acquisition of an irregular past tense form is a process of competition between a certain irregular phonological rule (e.g. *-t* suffixation & vowel shortening as in *lose, deal*) and the default *-ed* rule, where overregularization errors result from failures to apply appropriate irregular phonological rules over the default rule. Importantly, in the RC model, the performance of an irregular verb is determined by two factors: the correct identification of class membership and the probability of the irregular rule applying over the default *-ed* rule. Yang proposes the class-based frequency hierarchy: irrespective of the verb’s individual frequencies, which can be rather low, its correct usage rate can be quite high, as long as other members of its class are frequently encountered. In other words, high weight phonological rules enable low-frequency verbs to be used with high accuracy,

¹This idea is parallel to the pattern associator of Rumelhart and McClelland (1986), which was proposed to derive both regulars and irregulars by means of associative memory.

what Yang calls ‘the free-rider effect’. He claims, for example, that the verbs *hurt* and *cut* occur with high accuracy rate in children’s production despite their low absolute frequency, since they are in the same class as the verbs *hit*, *let*, *set*, *put*, etc., which have very high usage frequencies (totaling over 3000 occurrences), and every occurrence of such verbs increases the weight of the class rule.² Thus, the rule for this class, such as [- \emptyset & No Change], has a very high weight, which enables low-frequency verbs *hurt* and *cut* to be used with high accuracy. In contrast, *blew*, *grew*, *flew*, and *drew*, which belong to the class [- \emptyset & Rime \rightarrow u] are problematic, since these verbs are infrequent in the same way as the other verbs in this class, such as *know*, *throw*, etc. As Yang points out, the [- \emptyset & Rime \rightarrow u] class totals only 125 occurrences in the input sample. Hence the weight of the rule [- \emptyset & Rime \rightarrow u] is lower than that of the rule [- \emptyset & No Change], which explains the asymmetry in children’s performance with these verb classes reflected in the accuracy rates.

In sum, both the WR and the RC models focus on children’s overgeneralization patterns in order to explain the asymmetries in children’s production of past tense forms of irregular verbs. Within the WR model token frequency is claimed to have an effect on the rate of acquisition, hence rote-based learning must take place. Importantly, frequency effects are only predicted for irregulars, regular forms are acquired by the rule. Within the RC model, however, the differences in children’s production for individual verbs are explained in terms of their class membership, so that their individual frequencies have no effect. Hence it is a rule-based process from the start.

4.2.3 Rule- vs. rote-based gender acquisition?

Following the ideas developed within the WR and the RC model, I suggest that the examination of children’s overgeneralization errors can shed light on whether the hypothesized formal-to-semantic transition is abrupt or gradual, and more specifically, whether the token frequency of individual nouns has an impact on the acquisition process. Studying the (non-)effects of frequency on children’s performance can thus help answering the question whether gender acquisition is a rule-based process from the start or whether rote-based learning is also involved. This examination also allows me to explore the implications of novel data for the WR and the RC models.

²The child data used by Yang (2002) comes from Marcus (1992) where four American children were studied (Adam 2;3-5;2, Eve, 1;6-2;3, Sarah 2;3-5;1, and Abe 2;5-5;0). In addition Yang analyzed the input sample of more than 110,000 adult sentences to which these children were exposed during the recordings.

To achieve these goals I examine children’s production with male kinship terms, such as *papa* ‘daddy’, *deduška* ‘grandad’, *djadja* ‘uncle/man’, *mužčina* ‘man’, and *junoša* ‘youth’. In terms of frequency, these nouns should be different: *papa*, *deduška*, and *djadja* are expected to occur rather frequently in child directed speech, while *mužčina* and *junoša* must be rare. The frequencies discussed below were attested in a sample of the mother’s speech from Protassova’s corpus, CHILDES Database (MacWhinney 2000). The child Varvara, whose mother’s utterances I analyzed, was aged 1;6-2;10. The adult data taken for the analysis comprise eight files. The *kwal* command in the CLAN program was used to count the total number of tokens as well as the number of phrases containing agreement.³ The results are presented in Table 4.1. As expected, *papa* has the highest frequency in this sample of child directed speech. *Deduška* and *djadja* occur somewhat less frequently.⁴ In this sample of child-directed speech the noun *djadja* was used to denote a ‘man’ (i.e. not a relative) in all the occurrences. Finally, the nouns *mužčina* and *junoša* are not present in the adult speech at all.

Table 4.1: Frequency of occurrence of individual nouns and agreeing forms in the sample of child directed speech, MOT in eight files of Varvara (age 1;6-2;10)

Noun	N (nouns)	N (noun + Agr)
<i>papa</i> ‘daddy’	54	27
<i>deduška/deduka/deda</i> ‘grandad’	14	4
<i>djadja</i> ‘uncle/man’	5	1
<i>mužčina</i> ‘man’,	0	0
<i>junoša</i> ‘youth’	0	0

Admittedly, different children are exposed to different frequencies with these nouns, yet, it is the general pattern which is important here, i.e. *papa* ‘daddy’ is a high-frequency noun, *deduška* ‘grandad’ and *djadja* ‘uncle/man’ can be rather frequent, while *mužčina* ‘man’ and *junoša* ‘youth’ are rare.

With these results in mind the following predictions can be made. If the integration of the semantic principle is affected by token frequency and is

³The analysis command used for calculation was computed for each of the five nouns and contained the noun’s root, e.g. the formula ‘kwal@+s pap*+t*MOT’ was used to search for utterances with the noun *papa* ‘daddy’.

⁴Note that Russian families have different traditions to address a grandad. The child’s mother used three variants, *deduška/deduka/deda*, all three of them were counted in Table 4.1. The variant *deduška* has been chosen in this study in order to achieve a balance at the levels of syllabic and phonological structure with the other words in this class, such as *mužčina* and *junoša*.

thus an initially rote-based process, as we may predict from the WR model, there should be differences in children's agreement production for the individual *papa*-type nouns. More specifically, low-frequency nouns should exhibit higher rate of overgeneralization errors as compared to the other nouns in this class. In the RC model it can be predicted that the low-frequency nouns *mužčina* 'man' and *junoša* 'youth' should get a 'free ride' from the high frequency nouns in the same class, such as *papa* 'daddy'. Therefore there should be no frequency-overgeneralization correlation for the individual nouns within this class.

4.3 Acquisition of variable forms

Recall from Chapter 2 Sections 2.4.3 and 2.4.4 that in the case of hybrids and female names in *-ok/-ik* the semantic rule is used variably but predominantly in adult language and that its likelihood is very high in the case of verbal predicates. Based on these facts we can expect children to either match the adults or to be different. In any case children's behavior should be compared to that of their primary caregivers (parents, grandparents or older siblings). Since under natural conditions the situation when these nouns refer to a female might occur only rarely (not to mention the problem of time and contextual interpretation) it was decided to obtain adults' data through the same experimental testing.⁵

Furthermore, recall the finding from previous acquisition research on children's acquisition of variable input, namely that the acquisition of sociolinguistic constraints on variability can be delayed as compared to the grammatical constraints. Given that for hybrids and female names in *-ok/-ik* the use of the semantic rule is socially motivated (i.e. it emphasizes that it is a female that is engaged in a certain occupation) and that alternative syntactic agreement is also possible, the semantic rule can be acquired late. This means that children's agreement production for these nouns will not match the pattern of their caregivers, who are expected to produce semantic agreement most of the time with verbs. This finding also suggests that in variable contexts, where the semantic rule has a socio-cultural status, it will not be acquired simultaneously with the semantic rule in obligatory contexts, where it has categorical status. This means that the course of acquisition of the semantic rule in variable vs. obligatory contexts will be different.

⁵For example, in the sample of Varvara's mother's speech discussed in the previous section, which I searched for various professional titles, I only found three occurrences of the noun *doktor* 'doctor', all of which were used in the present tense. This means that they contained no indication of gender agreement.

Finally, one cannot exclude the possibility that these socio-cultural constraints are irrelevant for children, hence the following prediction could be an alternative: as soon as children realize the grammatical function of the semantic principle with one noun class they can automatically extend this knowledge to other noun classes.

4.4 Referential gender

As described in Chapter 2, the gender of double gender nouns, which have feminine morphology, is established referentially on each occasion, i.e. the same noun is assigned masculine if the referent is a male and feminine, if it is a female. If children do not know that reference is the determinant factor, they may decide that double gender nouns are just feminine as suggested by their form. This implies that they will not be willing to use the semantic principle for them, even when shown a picture of a female. This of course can only be seen when there is a mismatch between the two gender criteria, i.e. when a double gender noun refers to a male.

From the anecdotal evidence mentioned in Chapter 3 it also appears that children who are aware of form-gender correspondences may be uncomfortable to use a ‘feminine looking’ noun for a male, thereby they can adjust the noun shape in order to match the properties of the individual it describes (cf. (20) in Section 3.3). Given this, it is to be expected that the acquisition of the semantic principle for double gender nouns referring to males will be late acquired, since morphology can be an interfering factor. Therefore we can expect to find an asymmetry in children’s production with higher accuracy rates for females than for males.

4.5 Classes of nouns

The first of the two alternative predictions that can be made regarding the acquisition of the semantic rule across the classes of nouns is the following: if the children realize the dominance of the semantic rule for one subcategory of nouns they may do so uniformly and simultaneously for all nouns. In other words, if children reveal the knowledge of the semantic principle in one grammatical domain, they should be able to apply it across-the-board.

An alternative prediction is related to the idea expressed recently by Roeper (2007). Based on the evidence from various acquisition areas (e.g. Verb Second, definite articles, English past tense verb formation), he proposed that children distinguish classes of categories, i.e. they pay attention

to the notion of class in terms of structural as well as semantic and pragmatic distinctions, and generalize within it. According to Roeper, in order to be able to extend the knowledge of a particular language phenomenon to a new domain, the child should receive sufficient evidence for it in the input.

On this view, I do not expect the semantic rule to be acquired simultaneously for all nouns. Specifically, I can expect that female names in *-ok/-ik* and especially hybrids may pose a problem for gender learners due to the inconsistency/variability in the input. It can also be predicted that the referential gender of double gender nouns may not be easily acquired for several reasons. First, a sex-distinction is not part of their lexical entry. This means that it cannot be acquired once and for all, but must be established on every occurrence of a noun in a concrete discourse situation. Second, despite the consistency in terms of reference, two agreement patterns may be available for double gender nouns in the input. In fact, children may experience one noun to occur only with feminine agreement, another only with masculine, and still others with both. Based on these assumptions it can be predicted that the semantic principle will not be acquired uniformly and simultaneously for the all noun classes I am looking at. More specifically, it can be delayed for female names, hybrids and double gender nouns.

4.6 Summary

The predictions discussed in this chapter address various important issues, such as frequency, variation in the input, and referential properties of nouns, each of which is related to a particular subcategory of nouns. These predictions lay the groundwork for my experimentation and will be formulated in a more specific way at the beginning of each chapter of analysis.

The predictions regarding the acquisition of gender across the noun classes are aimed to generalize the pattern of acquisition of the semantic principle; they will be kept in mind throughout the chapters of analysis and brought into focus in Chapter 9 where the main results of the study are summarized and reconsidered.

However, before we can consider the results, it is necessary to discuss the methodology used in this experimental study. This is the topic of the next chapter.

Chapter 5

Methodology

5.1 Introduction

This study takes an experimental approach for assessing children's linguistic competence of gender. The present chapter provides a general description of the children that participated in the experimentation as well as a specification of the methodology used in the study. The goal of the chapter is to present the research strategy, i.e. the elicited production technique, and the design of eight experimental tasks. In Section 5.2 I introduce the participants of the study. The method is presented in Section 5.3 and the data collection process is described in Section 5.4. Section 5.5 discusses relevant transcription issues. Finally, in sections 5.6 to 5.13 I describe the goals, materials and procedure of each experimental task individually.

5.2 Description of the participants

The data used for the present study were collected at day-care center *Detskij Mir* in Ivanovo, Russia, in the spring of 2006.¹ Twenty-five normally developing children (14 girls and 11 boys) between the ages of 2;6 and 4;0 took part in the study. The complete list of the participants including their age is given in Table 10.1 in Appendix I. All children were monolingual speakers of Russian. Originally, the group of participants was slightly larger (29 children), but later it reduced in number due to various reasons. For example, some children lost interest in the game or got sick and could not complete the whole set of tasks.

¹Pseudonyms were assigned to the day-care center, as well as all the participants in the study, to protect their privacy.

The children's primary caregivers - mainly mothers, but also fathers, grandparents or elder siblings - were also asked to perform three tasks, i.e. Experiments 4, 5, and 6. This was done in order to compare children's and adults' agreement strategies in variable contexts. In the course of experimentation it was also decided to carry out these tasks with older children in order to explore further development in the acquisition of the semantic principle in variable contexts. For this purpose I recruited 12 children (four boys and eight girls) between the ages of 5;1 and 6;5 who attended the same day-care center.

5.3 Elicited production

In the study I chose to adopt an Elicited Production Task. This technique is widely used in experimental research on first language acquisition in order to access children's linguistic competence (cf. Thornton 1996, Crain and Thornton 1998). Specifically, it is designed to reveal children's grammars by having them produce particular sentence structures. As Crain and Thornton (1998) point out, both children's adult-like production and most interestingly their non-adult responses are of great importance, as they can provide insights into children's grammars and into the nature of the acquisition process itself. The structures of interest are usually elicited in the broader context of a game, in which the experimenter provides the context for production of a certain structure without modeling it. This means that the experimenter's 'lead-in' statement must not reveal the grammatical properties of the target structure. According to Thornton (1996), the technique of elicited production can be successfully used with children from the age of 2;6.

In this study the elicited production method is particularly appropriate for several reasons. First, the focus of the study is on specific noun types, which might appear rarely in young children's spontaneous speech. For this reason it was necessary to create a felicitous environment for agreement constructions with particular nouns to occur. The elicited production technique thus allowed me to explore the specific phenomenon of interest in a controlled fashion, i.e. it allowed me to put children in a situation that induced them to use agreement with the target nouns. Second, by using this technique I was able to elicit gender agreement for invented novel nouns that do not exist in the lexicon and thus cannot be familiar to children from their everyday language experience. Third, I was able to elicit data from children of various ages, which allowed me to draw conclusions about their grammar at particular points in time. Finally, this technique allowed me to gather a relatively large data sample within a short period of two months.

For the purposes of the study, eight elicited production experiments were designed to cover various noun types and to address the specific hypotheses. The target structures were elicited by devising situations that were uniquely felicitous for a specific noun group as well as for the specific agreement type. In all the tasks the children were involved in a game in which they interacted with the experimenter, i.e. the author of the dissertation. No assistant was engaged in any of the tasks. Sometimes the experimenter played the role of a hand puppet called Elmo. This was done in order to make the sessions more enjoyable for the children and to keep their attention focused on the task. It should be noted that the children were generally willing to interact with the experimenter as much as with the puppet.

The experiments were designed by the author and piloted with three Russian children aged 3 and 4 who lived in Tromsø in 2005 and 2006. These three children attended day-care centers where they spoke Norwegian, but they had monolingual Russian parents and spoke only Russian in the family. The experiments were also piloted with their parents.

Finally, Experiments 1, 2, and 3 were also piloted with 15 monolingual Russian children (five girls and ten boys) between the ages of 2;3 and 4;2. These tasks were carried out at the day-care center *Detstvo* in Ivanovo, Russia in the spring 2005, i.e. a year before the main study was conducted. Pilot testing allowed me to improve the design and the test condition in the main study. The changes will be discussed in Sections 5.6, 5.7, and 5.8 with regard to each task. This is important, since in Chapter 6 the results of the main study will be compared to the results of the pilot test.

5.4 Data collection

As pointed out by various researchers (Crain and Thornton 1998, Schaeffer 2000 *inter alia*), it is very important to establish contact with children prior to testing in order to achieve an optimal level of interest and comfort for them while performing the task. Therefore, several days before I started my experimentation I got acquainted with every child and spent time with them in the day-care center, playing and doing their every-day activities. This way, I got to know the children and made them familiar with me and the puppet. Later it was explained to the children that the puppet brought many interesting books and games with him, but he did not want to read and play alone. The children were also told that the puppet wanted to play with each of them individually. When they were asked whether they wanted to participate almost all the children volunteered. Those children who were somewhat skeptical at the beginning felt more relaxed and eager to play

during later sessions.

Each child was tested individually in a separate quiet room at the day care center. Usually it was a bedroom where a small table (that faced the wall) and two chairs were put during the sessions. This way, the children were familiar with the place and did not get distracted as there were only beds in that room. The children were tested every day. Each child was tested for a period of almost three weeks. Importantly, the experimentation was carried out twice. First for 13 children aged 2;6-3;3, who attended the same age group in the day-care center, and later for 12 children aged 3;6-4;0, who attended the other age group. This was done in order to minimize the likelihood of a developmental change in the children's grammar. Each group was tested for a period of almost three weeks. The sessions were recorded with a lightweight minidisk recorder (Sony MZ-R70) by using a stationary microphone (Sony ECM-MS907). Each recording session lasted for about 10-15 minutes with each child. If the child was tired or not interested in performing the task the recording was stopped and finished later. The children were recorded in the first part of the day before lunch and in the second part of the day after they had slept.

The experiments were carried out in the order that they are described in Sections 5.6 to 5.13. This is especially important for the experiments 5, 6, 7, and 8, where hybrids and double gender nouns were tested in two types of contexts. The tasks where referents of the nouns were absent were performed before those where they were present, so that children would not be biased in their agreement choices by the sex of the referents they already saw in the previous stories.

As I said in Section 5.3, the data were elicited in the context of a game. Specifically, the first task (i.e. Experiment 1) was organized as a puzzle game, where the children and the puppet played with characters of different colors that represented the test nouns. The other experiments were storybook reading tasks where children were asked to complete the stories told by the puppet or the experimenter in the presence of picture contexts. During the experimentation I tried to use exactly the same lead-in procedure for all the children.

The three experiments which were carried out also on adults were identical to the experiments carried out with children. This was done in order to avoid the discrepancy in methodology. Adults were tested individually in a separate room at the day-care center. These tasks are presented in Sections 5.9, 5.10, and 5.11.

Finally, the materials used in the experimentation were produced by the author. Most of the pictures were downloaded from <http://office.microsoft.com/en-us/clipart/default.aspx> and modified according to the experimental

design. The pictures were printed in color and laminated.

5.5 Transcription

The recordings were transcribed shortly after the experimentation was completed. Transcription was carried out by the author (I occasionally consulted Eugenia Romanova, another native speaker of Russian). Excel sheets were used to fill in the data. The data were grouped, so that one sheet covered one experiment and included the transcribed responses from all 25 participants. The data from caregivers and older children were organized in the same way. The transcripts were not coded. They were transliterated in Latin orthography and illustrate mainly syntactic and morphological peculiarities. There is no specification of phonology or intonation. When examples from the child data are used in the discussion in the following chapters, they are presented in the same form as they have been transcribed. Table 5.1 illustrates the relevant symbols from CHAT notation that were used in the transcription.² The examples from the child data presented in Chapters 6-8 will occasionally include some of these symbols.

Table 5.1: Overview of the symbols used in transcription

Symbol	Meaning
xx	unintelligible word
+	incomplete word
#	pause
[?]	uncertainty on the part of the transcriber of material in the angled brackets
+/	interruption
+//	self-interruption

5.6 Experiment 1: *Papa*-type nouns

5.6.1 Goal of the experiment

The experiment was designed to elicit adjectival agreement (e.g. *sinij papa* ‘blue father’ or *papa sinij* ‘father is blue’) as well as verbal predicate agreement (e.g. *papa upal* ‘father fell down’) with five male kinship terms: *papa* ‘fa-

²CHAT is the Codes for the Human Analysis of Transcripts available at <http://chilides.psy.cmu.edu/manuals/chat.pdf>.

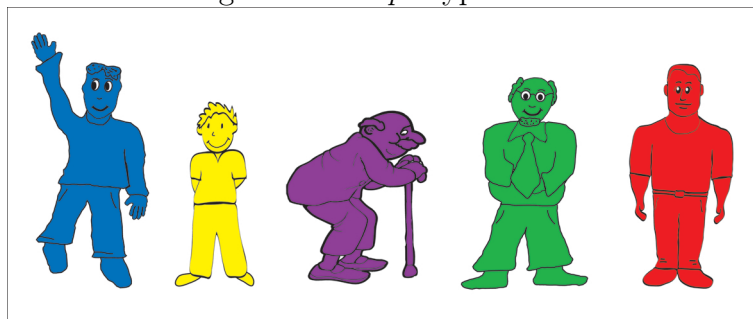
ther', *deduška* 'grandfather', *djadja* 'uncle/man', *mužčina* 'man' and *junoša* 'youth'. Other masculine, feminine and neuter nouns were used as fillers, e.g. *lev*_(M) 'lion', *pingvin*_(M) 'penguin', *cyplenok*_(M) 'chicken', *slon*_(M) 'elephant', *mama*_(F) 'mother', *kurica*_(F) 'hen', *sova*_(F) 'owl', *čerepaxa*_(F) 'turtle', *kolesó*_(N) 'wheel', *peró*_(N) 'feather', *pomeló*_(N) 'broom', *vedró*_(N) 'bucket', *kol'có*_(N) 'ring'. All filler items had a transparent morphological form, i.e. the masculine nouns ended in a hard consonant, feminine nouns ended in *-a*, and the neuter nouns ended in a stressed *-o*. All nouns were presented in nominative singular.

In the pilot test I elicited adjectival agreement only and tested four nouns: *papa* 'father', *djadja* 'uncle/man', *mužčina* 'man' and *junoša* 'youth'. The noun *deduška* 'grandfather', which as I mentioned in Chapter 4 has several variants, was added to the main study in order to achieve balance in terms of syllabic structure and sound shape between the short familiar nouns and the long rare nouns, i.e. *mužčina* 'man' and *junoša* 'youth'. In other words, this was done in order to exclude the possibility that the phonological complexity of a noun could affect the performance for familiar vs. rare items. Finally, since verbal agreement was elicited in Experiments 2 and 3, it was decided to change the design of this experiment in the main study and elicit noun-verb agreement in addition to the adjective-noun forms.

5.6.2 Materials and procedure

The experiment was introduced as a game where cardboard characters of different colors were used to represent each noun. Each character appeared in five colors: blue, yellow, red, green and purple. Hence there were five fathers, five grandfathers, etc. The characters representing the test nouns *papa* 'father', *junoša* 'youth', *deduška* 'grandfather', *djadja* 'uncle/man', and *mužčina* 'man' are illustrated in Figure 5.1.

Figure 5.1: *Papa*-type nouns



Every test item was introduced in a separate trial together with three

fillers: masculine, feminine, and neuter. Thus, there were five trials performed on five different days. The characters representing each noun were placed into small paper bags which were put on the table as well as some small objects, e.g. a book, a saucer, a cup, etc.

In the introduction I explained to the children that Elmo was a silly puppet who could not remember the names of colors and who refused to listen to adults. The child was then asked to help Elmo learn the color terms. After that I explained the rules of the game. In this task the experimenter was taking the characters out of the bags and put them in different places, e.g. under the book. The child had to tell Elmo what color character was where, e.g. *sinij papa pod knjigoj* ‘blue father is under the book’. During the trial session different color eggs were used to explain the game. The neuter noun *jajc6* ‘egg’ was used in the introduction, so that it would not let the children guess about the agreement with the test nouns. As soon as I got the impression that the child understood the task I introduced the characters which were all white and asked the child to repeat their names. Then I asked the child whether s/he wanted to see what color characters were in the bags and teach Elmo those colors. When the characters were taken out of the bags I used the following lead-in statement: *Posmotri, vot papa. A po cvetu papa?* [Here is a father. And what color is father?³] The character was then placed on the table and I asked: *Skaži Elmo gde teper’ papa* [Tell Elmo where father is now.] If the child forgot to use the color term I reminded that it was important to name the color of each character; otherwise Elmo would get mixed up. If the child used wrong color adjective s/he was never corrected. Importantly, I controlled the order of colors and characters, so that the test item, i.e. a male kinship term, always followed after a neuter filler of a different color. This was done in order to avoid a “carryover” effect, i.e. the similarity of materials across experimental conditions.

5.7 Experiment 2: Male names in -a

5.7.1 Goal of the experiment

This experiment aimed at eliciting predicate agreement with hypochoristic forms of male names in -a and past tense of the verb. In the pilot test this

³Russian variant of this question used in the task contains an adverb *po cvetu* ‘by color’. It would be more natural to ask *Kakogo papa cveta?* ‘What_{GEN.M} color is father?’ However, this lead-in question was tried in the pilot test and was rejected, since this question presupposes the answer *zelenogo* ‘green_{GEN.M}’, where the adjective agrees with the noun *cvet* ‘color_M’, but not with the test noun, which would be *zelenyj* ‘green_{NOM.M}’.

experiment was designed to test modern familiar names. Therefore there was a possibility that children could know the gender of these nouns from before or they could simply associate these familiar names with concrete individuals. In order to exclude the familiarity effect, very old-fashioned, rare names were used in the main study, e.g. *Trenya*. This was done in order to exclude the possibility that children could know the gender of these nouns from before. Rare full male names (e.g. *Agap*), whose morphological form corresponds to their gender as well as female names (e.g. *Luša*) were used in the task as filler items. The complete list of names is given in Table 5.2.

Table 5.2: The list of proper names from Experiment 2

male names		female names
short	full	
Mulya	Agap	Luša
Mikeša	Fedot	Fekla
Troša	Jasik	Glaša
Ganya	Osip	Nyura
Trenya	Trofim	Bronya
Jan'ka	Polivan	Agaf'ya
Vaxonya	Siluyan	
Lutoxa	Agafon	
Ošanya		
Foma		

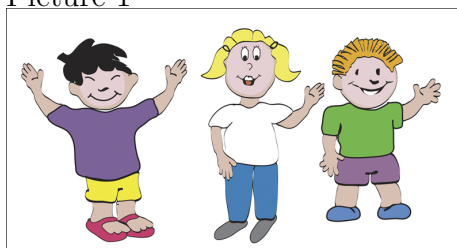
5.7.2 Materials and procedure

A set of pictures was made for the experiment. Each picture portrayed three children: two boys and a girl, or three boys. The pictures were paired: the first one introduced the characters, as in Figure 5.2, and in the next picture they were shown such that they had performed some action, as in Figure 5.3.

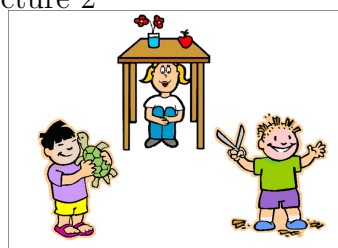
The children were first shown the characters in the first picture and heard their names. After that they repeated the names together with the experimenter. The children were also asked to call the characters by names. There was no training session in this task. Next they were asked to say what each child had done in the picture that followed. The experimenter pointed to a single child, usually starting with the test item, and asked the following lead-in question: *Čto *sdelali_{PL} (target: sdelal_{M.SG}) Trenya?* [What did *Trenya* do?]. Note that the lead-in question is ungrammatical, since the verb has the plural form. This was done in order to avoid producing the target structure,

Figure 5.2: Male names in *-a*:

Picture 1

Figure 5.3: Male names in *-a*:

Picture 2



which reveals the agreement. The technique was inspired by Popova (1973) who used the same question form in her experimental design.⁴

5.8 Experiment 3: Novel nouns

5.8.1 Goal of the experiment

In this storybook reading task I introduced the children to a novel (invented) noun *obormoša*. The noun was used to denote a non-existing animal of male gender. The sex of the character was explained to the children in the introduction. I have also tried to portray the animal ‘male looking’. Thus, this noun showed a form-meaning mismatch similar to that of *papa*-type nouns: its ending is typical of feminine nouns but semantically it should be masculine. It should be noted that in the pilot test a different novel noun was used, namely *čunja*. In the course of experimentation I realized that the disyllabic structure of this noun made it similar to male names in *-a*, which are mainly disyllabic. Therefore, it was decided to use a different non-existing noun in the main study, i.e. *obormoša*. The length and the sound-complexity of the noun was chosen so that it would be approximately as hard as for the existing nouns *mužčina* ‘man’ and *junoša* ‘youth’. The experiment aimed at eliciting attributive adjective agreement as well as verbal predicate agreement.

⁴The technique proved to be successful, i.e. in general the children were not affected by the ungrammatical plural agreement in the lead-in question. I found only eight plural verb forms in the speech of four children (one child made four such errors in different tasks and other children made one error each). These eight errors occurred in five tasks (Experiment 2, 3, 5, 6, 8). However, in Experiment 7, where double gender nouns were used non-generically while their referents’ biological sex was unknown, 28 (9.9%) plural verb forms were found in the speech on 11 children. This result will be explained in Chapter 8.

5.8.2 Materials and procedure

A set of pictures representing ten identical but differently colored, but otherwise identical imaginary animals called *obormoša* was used in the task. The first picture showed a single *obormoša*, as shown in Figure 5.4. The child was told that this was a boy and was asked to pronounce the name several times. The child was also told that the animal liked to be called by name very much. Then the child was invited to listen to the story about ten different-color *obormoša*. The story was told by the experimenter who was manipulating the puppet called Elmo. Elmo was said to be a little lazy and forgetful, and the child was asked to help him tell the story.

Figure 5.4: *Obormoša*: Picture 1

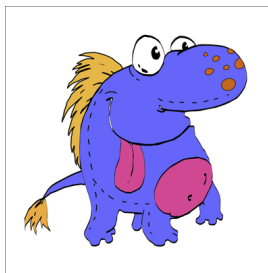
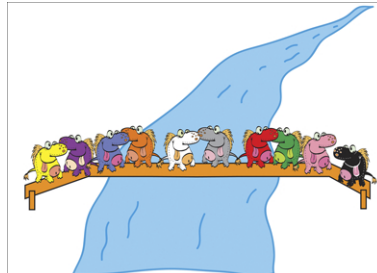


Figure 5.5: *Obormoša*: Picture 2



Then followed the picture in which ten animals were sitting on the bridge, as in Figure 5.5, and the puppet said:

Posmotri, vot desjat' raznocvetnyx obormoš. Davaj posčitaem ix vmeste.

[Look, there are ten different-color animals. Let us count them together.] (Then the child was counting together with the puppet.)

A ty znaeš kakix oni cvetov? V ètoj istorii očen' važno različat' cveta i nazyvaj' každoe životnoe po cvetu, čtoby obormoši ne pereputališ'.

[Do you know what colors they are? In this story it is very important to distinguish the colors and to point out the color of each animal so that we know who is who.] (Then Elmo was pointing at a single *obormoša* and the child was providing a color for it. The content was never corrected.)

Kak-to raz obormošy sideli na mostike i ničego ne delali. I im stalo očen' skučno, i oni rešili poigrat'. Davaj posmotrim kak oni igrali?

[Once the animals were sitting on the bridge doing nothing. It was very boring and they decided to play. Let us see how they played.]

Then followed the picture in Figure 5.6. In the background, the child could see that only nine animals were still sitting on the bridge, and one was playing in the foreground. The puppet then asked a lead-in question: *Čto slučilo s obormošej?* [What happened to *obormoša*?] or *Čto *sdelali_{PL}* (target: *sdelal_{M.SG}*) *obormoša*? [What did *obormoša* do?] The former question contained neuter agreement, where the verb *slučilo* ‘happened’ agreed in neuter with the question word *čto* ‘what’. Thus agreement with the test noun was avoided. The latter question was in plural. By doing so, I carefully avoided providing any clues to the noun’s gender, except the animal’s name.

Figure 5.6: *Obormoša*: Picture 3



After the child answered the question, s/he helped Elmo to count how many animals were left on the bridge. The procedure continued till no animals were left on the bridge. If the child sometimes answered in present, the experimenter explained that the action has already happened by using the adverb *uže* ‘already’. After that the elicitation question was asked again. By doing so I tried to elicit as many utterances containing agreement as possible.

5.9 Experiment 4: Female names in -ok/-ik

5.9.1 Goal of the experiment

The goal of the experiment was to test children’s agreement strategies for nouns that yield variable agreement forms in adult grammar. More specifically, it aimed to examine children’s ability to use the semantic rule. The task was designed to elicit verbal predicate agreement, where the likelihood of semantic agreement on the verb is especially high in adult language (cf. the discussion in Chapter 2, Section 2.4.3). The following female names were used as test items: *Ninčik*, *Lenok*, *Valek*, *Marinčik*, *Dusik*, *Verok*, *Natusik*.

5.9.2 Materials and procedure

In this storybook reading task the children were asked to help the puppet tell stories about little girls. Six pictured stories were arranged into one book, each story contained several color pictures where girls were the main characters. In the introduction, the experimenter manipulating Elmo presented a girl in the picture and asked the child to repeat her name. Below I present the introduction to the story about two girls *Ninčik* and *Lenok*, which was accompanied by the picture in Figure 5.7:

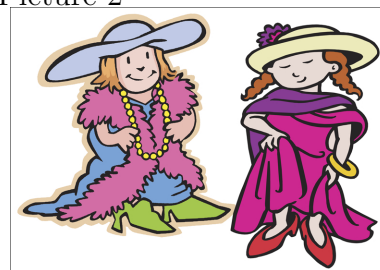
Ètix devoček zovut Lenok i Ninčik. Oni očen' neposlušnye. Kak-to raz ix mama pošla v magazin i ostavila ix doma odnix. Kogda mama ušla, Lenok i Ninčik otkryli bol'šoj škaf i stali nadevat' na sebja maminu odeždu. A v škafu byli plat'ja, šljapy, tufti na koblukax. Davaj teper' posmotrim, čto Lenok i Ninčik na sebja nadeli.

[These girls are Lenok and Ninčik. They are very naughty. Once their mother went shopping and left them at home alone. When mother went out, Lenok and Ninčik opened a big wardrobe and started trying the mother's clothes on. There were dresses, hats, and shoes on high heels in the wardrobe. Let us see now what Lenok and Ninčik put on.]

Figure 5.7: Lenok and Ninčik:
Picture 1



Figure 5.8: Lenok and Ninčik:
Picture 2



In the preamble the experimenter carefully avoided using agreement with the test items. The lead-in question: *Čto Lenok i Ninčik nadeli?* [What did Lenok i Ninčik put on?], was accompanied by the picture in Figure 5.8. When answering the question the child was looking at the picture of a single child and the experimenter said: *Snačala Lenok... A teper' Ninčik...* [First, Lenok... And now Ninčik...]

5.10 Experiment 5: Hybrid nouns - referent absent

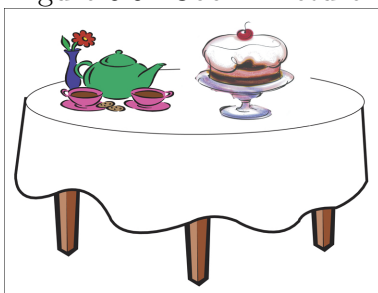
5.10.1 Goal of the experiment

The experiment aimed to elicit verbal predicate agreement with four hybrid nouns: *maljar* ‘painter’, *sadovnik* ‘gardener’, *povar* ‘cook’, and *dvornik* ‘yard keeper’. Importantly, in this experimental condition the sex of the referents was inaccessible, so that the noun’s morphology typical of masculine nouns was the only clue provided by the experimenter.

5.10.2 Materials and procedure

Four pictured stories (according to the number of the test nouns) were arranged in one book. Each story showed several color pictures of objects; no characters were present in any of them. In the preamble the experimenter explained that the stories in the book were about people of different professions who knew how to do various things. The experimenter also said that the child would not see those people in the pictures, because they finished their work and went to have a rest. Then the child was invited to see what had been done. Below I illustrate the story about a cook, which had the following introduction accompanied by the picture in Figure 5.9 (this picture was one out of three for the noun *povar* ‘cook’):

Figure 5.9: Cook: Picture 1



Èta istorija pro povara. Ty navernoè znaeš, èto povar umeet xorošo gotovit'. Odnadždy povaru bylo dano zadanie nakryt' stol k čajju: ispeč' tort, sdelat' čaj, postavit' na stol čaški i bljudca. Davaj teper' posmotrim, kak èto vse u povara polučilos'.

[This story is about a cook. Perhaps you know that a cook can make food very well. Once a cook was asked to make the table

ready for tea: to bake a cake, to make tea, to put cups and saucers on the table. Let us see now how the cook managed all this.]

In the preamble the experimenter carefully avoided producing agreement with the test item by using present tense or passive constructions. The following lead-in statement was used in the task: *Posmotri, povara uže net, no vse gotovo. Čto že bylo sdelano povarom? Čto *sdelali_{PL} (target: sdelal_{M.SG}) povar? Snačala, . . .* [Look, the cook is already gone, but everything is ready. What has been done by the cook? First, . . .] The story continued and other pictures were shown to the child.

5.11 Experiment 6: Hybrid nouns - referent present

5.11.1 Goal of the experiment

In this storybook reading task I aimed to test children's agreement strategies for hybrid nouns whose referents were females. The experimental items included seven nouns, such as *počta' on* 'postwoman', *doktor* 'doctor', *milicioner* 'policewoman', *fotograf* 'photographer', *povar* 'cook', *vrač* 'physician', *bibliotekar'* 'librarian'. In order to neutralize the pattern, I also included four hybrid nouns whose referents were males, such as *milicioner* 'policeman', *šofer* 'driver', *sadovnik* 'gardener', *povar* 'cook'. As in Experiment 4, I aimed at eliciting noun-verb forms in past tense in order to increase the likelihood of semantic agreement in children's production.

5.11.2 Materials and procedure

Two books containing six and five stories each (according to the total number of test items) were used in the task. Each story had several color pictures which portrayed female or male individuals of different professions, as well as other characters (usually children). The stories about males were used as filler items among the stories about females. The experiment was organized into two trials which were conducted on two different days. One book was read on each trial. Below I illustrate the story about a postwoman which had the following introduction accompanied by the picture in Figure 5.10:

Èta istorija pro devočku Mašu, ee babušku i tetju počta' ona. Kako raz pod Novyj God Maša zaboleda. Togda babuška rešila ee poradovat' i kupila ej podarok. No žila ona očen' daleko, v drugom gorode, poetomu ona poslala podarok po počte. Na drugoj

den' babuška pozvonila Maše i govorit: "Mašen'ka naverno segodnja priđet počtal'on i prineset tebe moj podarok". Maša očer' obradovalas' i stala ždat'. Vdrug v dver' pozvonili. Maša podumala, čto èto počtal'on i otkryla dver'.

[This story is about a girl called Maša, her granny and a post woman. Once around New Year's Day Maša got sick. So her granny decided to cheer her up and bought her a present. But she lived far away in another city, so she sent her present in the mail. The next day she called Maša and said: "Maša, I think today *počtal'on* will come and will bring you my present." Maša became very glad and she started waiting. Suddenly the door bell rang. Maša thought it was *počtal'on* and opened the door.]

Figure 5.10: Post woman: Picture 1



The following pictures were presented to the children one after the other.

Figure 5.11: Post woman: Picture 2



Figure 5.12: Post woman: Picture 3



While showing the picture in Figure 5.11 the experimenter produced the lead-in statement:

*Posmotri, vot zdes' Maša, a vot zdes' počtal'on. Ty konečno pomniš, čto počtal'on v ètoj istorii teten'ka. Vidiš, u počtal'ona bo'šaja sumka s piš'mami i podrok. A čto bylo da'še, kak ty dumaeš? Čto *sdelali_{PL} (target: sdelal_{M.SG}) počtal'on?*

[Look, here is Maša and here is *počtal'on*. You surely remember that *počtal'on* is a woman in this story. Look, the post person has a big bag with letters and a present. What do you think happened then? What did *počtal'on* do?]

The picture in Figure 5.12 was used to help children with the answer. Note that in every story the sex of the character representing the test noun was made salient in the drawing as well as orally.

5.12 Experiment 7: Double gender nouns - referent absent

5.12.1 Goal of the experiment

The experiment was designed to elicit verbal predicate agreement with five double gender nouns which were derived from real words (e.g. *narjažalka* ‘decorator’ from the verb *narjažat'* ‘to decorate’) with some typical for this type of nouns suffixes, e.g. *-ka*, *-xa*, etc. These nouns are *narjažalka* ‘decorator’, *žalalka* ‘pitiful person’, *umnjaša* ‘smarty pants’, *pomoguša* ‘helper’, and *pačkuxa* ‘sloven’. These nouns are thus made-up but not novel, as e.g. *obormoša*, which was used in Experiment 3 (cf. Section 5.8). In this experimental condition the sex of the referents was unaccessible, so that the nouns’ morphology typical of feminine nouns was the only clue provided by the experimenter.

5.12.2 Materials and procedure

Five pictured stories (according to the number of test nouns) were arranged in one book. No characters were shown in any of the pictures. In the preamble the experimenter explained that the stories in the book were about children. The experimenter also said that there would be no children in the pictures, because they did something and left. Then the child was invited to see what had been done. Below I illustrate the story about a child called *narjažalka* ‘decorator’, which had the following introduction accompanied by the picture in Figure 5.13:

Èta istorija pro narjažalku. narjažalka ljubiti vse narjažati, čtoby bylo krasivo. Poètomu kak-to raz, narjažalku poprosili ukrašiti vot takuju eločku k Novomu Godu: povesiti na nee raznocvetnye šariki, a na makušku nadeť zvezdu. Davaj teper' posmotrim, čto bylo daľše. Čto bylo sdelano narjažalkoj?

[This story is about *narjažalka*. *narjažalka* likes to decorate everything, so that it looks nice around. Therefore, *narjažalka* was asked to decorate this tree for the New Year's Day: to hang different color balls, and to put a star on the top. Let us see now what happened afterwards. What was done by *narjažalka*?]

Figure 5.13: Decorator: Picture 1



The other two pictures presented in Figures 5.14 and 5.15 were shown one by one and the experimenter asked: *Čto bylo sdeleno narajažalkoj? Čto *sdelali_{PL} (target: sdela_{M.SG}) narjažalka? Snačala . . . A čto bylo sdeleno potom?* [What was done by *narjažalka*? What did *narjažalka* do? First, . . . And what has been done after that?]

Figure 5.14: Decorator: Picture 2



Figure 5.15: Decorator: Picture 3



5.13 Experiment 8: Double gender nouns - referent present

5.13.1 Goal of the experiment

The experiment was designed to elicit verbal predicate agreement with four made-up double gender nouns, such as *umnjaša* 'smarty pants', *pačkuxa*

‘sloven’, *obižala* ‘bully’, *poedala* ‘heavy eater’, and two existing but presumably infrequent double gender nouns, viz. *stiljaga* ‘mod’ and *bedolaga* ‘poor wretch’. These nouns were used to refer to males. In addition, I used four made-up double gender nouns to refer to females, such as *umnjaša* ‘smarty pants’, *pačkuxa* ‘sloven’, *terjaxa* ‘flaky person’, *žalelka* ‘pitiful person’. These were filler items. As in the Experiment 4 and Experiment 6, I aimed to elicit noun-verb forms in past tense in order to increase the likelihood of semantic agreement in the children’s production.

5.13.2 Materials and procedure

Ten pictured stories (according to the number of test items) were arranged into two books. Each story was about a boy or a girl who were present in the pictures. The stories about girls were used as filler items among the stories about boys. The experiment was organized into two trials which were conducted on two different days. One book was read on each trial. Below I present the pictures and the introduction to the story about a boy called *bedolaga* ‘luckless man’.

Èta istorija pro bedolagu. V našej istorii bedolaga èto maľčik. Bedolaga vse vremja popadaet v neprijatnye istorii, poètomu tak i nazyvaetsja. Vot kakaja istorija priključilas' s bedolagoj na rečke.

[This story is about *bedolaga*. *Bedolaga* is a boy in our story. *Bedolaga* often gets into trouble, hence (he) is called this way. Here is a story that happened once by the river.]

Figure 5.16: Luckless man:
Picture 1

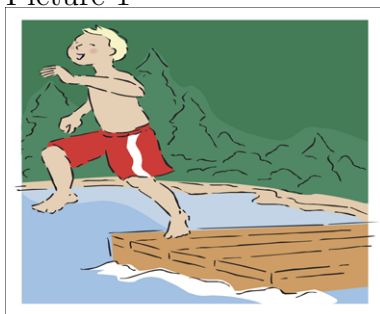
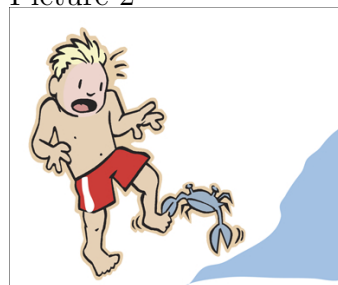


Figure 5.17: Luckless man:
Picture 2



The pictures in Figures 5.16 and 5.17 were shown one by one and the experimenter asked the lead-in question: *Čto slučilos' na rečke? Čto *sdelali_{PL}*

(tagret: *sdelal*_{M.SG}) *bedolaga*? [What happened by the river? What did *bedolaga* do?]

Note that in every story the sex of the character introducing the test noun was made salient in the drawing as well as orally.

5.14 Summary of the chapter

In this chapter I have specified the methodology used in the study. As we have seen, the elicited production experiments have been specifically designed to explore children's knowledge of the semantic principle with different sub-categories of nouns. Each experimental situation is uniquely felicitous for testing the predictions discussed in Chapter 4. In the following chapters, where the experimental results are presented and analyzed, the specific predictions are linked to the experiments and to the relevant noun classes. The results are presented in three chapters. Chapter 6 considers the acquisition of the semantic principle in obligatory contexts, i.e. for *papa*-type nouns and male names in *-a*. Chapter 7 focuses on the variable contexts regarding hybrids and female names in *-ok/-ik*. Finally, Chapter 8 targets the acquisition of referential gender with double gender nouns.

Chapter 6

Acquiring gender in obligatory contexts

6.1 Introduction

I begin my investigation by looking at nouns which are inherently specified for gender in the lexicon. These are common nouns like *papa* ‘daddy’ (also referred to as *papa*-type nouns) and male names in *-a* (e.g. *Vanya*) that have feminine case forms and take masculine agreement. In reality, these nouns belong to the masculine gender, since they refer to male individuals. Hence the dominance of the semantic rule for these nouns is a categorical requirement.

In Chapter 3 I have shown that the nouns in this class can be problematic for a Russian child between the ages of 2 and 3. Yet, as I said it is not possible to draw specific conclusions about the course of acquisition from the diary data of one child (Gvozdev 1961). Therefore the main aim of this study is to reveal a more general picture regarding the acquisition of the semantic principle with different classes of exceptional nouns, which may provide deeper insights into the course of acquisition and development of language.

The aim of this chapter is twofold: First, I present novel experimental results showing an asymmetry in agreement production for individual common nouns (*papa*-type nouns), as well as an asymmetry between rare proper names (male names in *-a*) and low-frequency common nouns. I then consider the implications of the new data for generative and cognitive-functional approaches to language acquisition, such as the Rules and Competition (RC) model (Yang 2002) and the Words and Rules (WR) model (Pinker 1999). Second, I explore the ramifications of these findings for understanding the

relationships between formal and non-formal mechanisms involved in the acquisition of gender.

The structure of this chapter is as follows. In Section 6.2 I formulate the specific hypotheses and predictions for the acquisition of gender with these nouns, some of which were previously discussed in Chapter 4. Other predictions are based on findings from Russian and other languages considered in Chapter 3, as well as certain theoretical considerations outlined in Chapter 2. In Section 6.3 I present the results of the Experiments 1, 2, and 3 and compare them to the results of the pilot test. Most crucially, both data sets show a contrast between high-frequency and low-frequency common nouns as well as between the rare male names in *-a* and low-frequency common nouns. In the analysis presented in Section 6.4 I propose that gender is a domain which is sensitive to token frequency. Nevertheless, I conclude that the results are not totally compatible with Pinker's WR model. The evidence here comes from children's agreement behavior with male names in *-a*. Furthermore, I suggest that there are other factors, which bear on the differences in the semantic representation of proper names vs. common nouns and which may explain why for some nouns the semantic rule is acquired more easily than for the others. In Section 6.4.4 I consider blocking principle as an explanation for the overgeneralization and the course of gender acquisition. The chapter ends with a summary of the conclusions (Section 6.5).

6.2 Hypotheses and predictions

Recall from Chapter 3 that the central question addressed in various acquisition studies has been whether children base their initial hypotheses on the formal properties of the nouns or on the semantic concepts related to natural gender. Based on data from French, German, Spanish, Czech, Russian, and Hebrew, it has been claimed that children's language organization changes from formal to non-formal, in the following way (cf. Karmiloff-Smith 1979, Mills 1986, Levy 1983b, Henzl 1975, Gvozdev 1961, Popova 1973):

- (1) The child starts out by making a formal grammatical analysis of the gender system. In the course of development formal rule-based generalizations are replaced by those made on the basis of the semantic properties of the nouns.

For Russian, the hypothesized change has been stated for the acquisition of *papa*-type nouns in both spontaneous (Gvozdev 1961) and experimental (Popova 1973) data. What remains unclear is how the acquisition of the semantic principle proceeds and what underlies this process. In terms of

gender assignment theory, we lack evidence showing how children establish the hierarchy of gender assignment rules, where a semantic rule (as in (4-a)) gains dominance over a morphological rule (as in (4-b)).

- (2) a. nouns denoting males are masculine;
 b. nouns of declensional class II are feminine.

As the acquisition strategy in (1) predicts, the establishment of gender for nouns like *papa* ‘daddy’ should take a longer time than for nouns that are not involved in gender conflicts. Recall from Chapter 3 that this delay has been pointed out by Gvozdev (1961). His son Zhenya, who started acquiring gender by paying attention to the morphological properties of nouns, was able to assign gender correctly to the nouns with transparent morphological form already by the age of 2;4. However, till approximately the age of 3;0 he could erroneously assign feminine gender to *papa*-type nouns and male names in *-a* that have feminine case forms, i.e. he occasionally produced feminine agreement for these nouns. According to Gvozdev, errors due to overgeneralization disappeared after age 3;0, which could be taken as an indication that the concept of natural gender is acquired by this time. This may be an important observation, which seems to be consistent with the claim that “...by three years normal children appear to complete the foundations of language acquisition” (Lust 2006:117). Based on Gvozdev’s findings it can be hypothesized that age 3 is an important point for the growth of gender knowledge in Russian children, the point when the semantic concept related to natural gender takes dominance over the morphological component. Yet, as mentioned before, Gvozdev’s findings are based on the diary study of a single child; therefore, additional evidence is needed to confirm this result. Therefore, it seems important to investigate children’s agreement production with the exceptional nouns before and after age 3 in order to establish the time when the semantic principle is acquired.

Before I formulate my other predictions I would like to define the term *acquisition* as it is used in the present dissertation. Traditionally, a grammatical structure has been considered acquired when it was used correctly in 90% of its obligatory contexts (cf. Brown 1973). The acquisition of gender in this dissertation is studied experimentally by using the elicited production technique, so that attributive adjective and verbal predicate agreement forms are the subject matter. Crain and Wexler (1999) proposed that in an experimental situation the researchers should expect a 90% accuracy level and up to 10% of responses being attributable to noise. This 10% or less includes performance errors, lack of attention, and noise of other kinds. Thus, the gender of a noun is not seen as acquired until the accuracy level of target-

consistent agreement production with this noun reaches 90%. However, this 90% criterion for use in obligatory contexts is irrelevant when alternative forms are available in the adult grammar, i.e. for the acquisition of hybrids and female names in *-ok/-ik*. In this case children are expected to perform in a target-consistent manner, i.e. the child's preferences should reflect those of adults.

With regard to my main predictions it seems especially important to examine children's agreement behavior, and in particular their overgeneralization errors, before age 3, when, as follows from Gvozdev's study, the development of the semantic principle takes place. Gvozdev's findings also suggest that the acquisition of the semantic concept may proceed slowly. In other words, the take-over described in (1) is not sudden, which corresponds to similar observations in other languages, e.g. in Czech, Hebrew, etc. (cf. the discussion in Chapter 3). On the other hand, as I said in Chapter 3, there is a broad consensus among researchers that formal gender features are acquired in a rule-based fashion (Karmiloff-Smith 1979, Mills 1986 *inter alia*). It has been shown that the fast establishment of the system of formal rules in a language is based on their consistency and clarity, rather than on frequency of individual nouns in the primary linguistic data. On this picture, an important question arises: Are formal and semantic principles acquired in the same fashion? The fact that the children gradually make use of semantic (sex-based) distinctions in gender acquisition may cast doubt on the existence of the rule-based learning mechanism here. To explore this issue one needs to find out whether frequency is involved in the acquisition of the semantic criterion.

Therefore it seems relevant to consider the acquisition of gender in light of the two acquisition theories: the Words and Rules (WR) model (Pinker 1999) and the Rule and Competition (RC) model (Yang 2002), which were reviewed in Chapter 4. According to these theories, we can expect the acquisition of the semantic principle to be either a dual or a single process. Along the lines of the WR model the semantic procedure should be a combination of both rote- and rule-learning. That is, the semantic principle should be acquired in several stages. First, children learn the gender of individual nouns by rote, and then they formulate the semantic rule and begin to generalize. Thus, the integration of the semantic rule takes time, since it involves some rote learning. In the RC model the semantic procedure is a rule from the start which competes for dominance with the morphological rule in the course of acquisition. In both models the gradualism of the acquisition process is attributed to input frequency. However, in the WR model it is token frequency, i.e. the frequency of occurrence of a particular noun in child-directed speech, that underlies gradual learning, while type frequency, i.e. the occurrence of

an item within a particular class, is argued to play a role in the RC model.

As discussed in Chapter 4, the class of *papa*-type nouns represents a good testing ground for both models, since on the one hand, it includes fairly rare items like *mužčina* ‘man’ and *junoša* ‘youth’, and on the other very frequent ones like *deduška* ‘grandad’, *djadja* ‘uncle/man’, and especially *papa* ‘daddy’. With regard to these nouns, the following frequency-overregularization correlation can be expected. First, if the semantic procedure involves item-based learning, as the WR model predicts, I should find higher accuracy rates for the nouns *papa*, *deduška*, and *djadja*, which occur in the input more frequently than *mužčina* and *junoša*. In other words, the overregularization errors, i.e. feminine agreement, should occur more often for *mužčina* and *junoša* than for *papa*, *deduška*, and *djadja*. Second, if the semantic procedure is a rule-based mechanism, as the RC model predicts, there should be no discrepancy in agreement production between high-frequency and low-frequency nouns within the same class, i.e. low-frequency nouns *mužčina* and *junoša* should occur with high accuracy rates, since they are in the same class with the high-frequency nouns *papa*, *deduška*, and *djadja*. Note also that this frequency-overregularization correlation may be most visible for two-year-olds, who according to Gvozdev’s study, may not have full mastery (at a 90% level) of the semantic principle yet.

Finally, to obtain a clearer picture on the nature of the semantic procedure, it seems necessary to compare children’s agreement behavior for familiar nouns with that of novel nouns. Therefore I included rare archaic male names in *-a* and a non-existing noun *obormoša* into the experimentation (see Chapter 5 Sections 5.7 and 5.8 for the description of the tasks). These nouns were used as a tool to assess children’s grammatical knowledge independently of their lexical knowledge (cf. Berko Gleason 1958). That is, if children can generalize the knowledge of natural gender to a word they have never heard in the input, this means that there must be a mental rule that allows them to do so. Children’s productive use of masculine agreement with novel and unfamiliar nouns should allow me to exclude the possibility that children’s gender knowledge is a result of memorization. The reverse result will indicate that children learn the gender of each lexical item one by one.

Hypotheses and predictions in sum:

Hypothesis I: Age 3;0 is a turning point in gender acquisition of *papa*-type nouns and male names in *-a*, which is associated with complete mastery (i.e. at the 90% level of all obligatory contexts) of the semantic principle.

Prediction 1: Although errors due to overgeneralization (i.e. feminine agreement forms) are expected in the speech of two-year-olds, children’s agreement production for these nouns should become target-consistent around age 3, so that after this period the error rates should not exceed the 10% of

the experimental error margin.¹

Hypothesis II: The acquisition of the semantic principle proceeds in an item-based fashion, which develops into a rule (cf. WR model, Pinker 1999).

Prediction 2: There should be differences in the accuracy rates between high-frequency nouns (*papa*, *deduška*, and *djadja*) and low-frequency nouns (*mužčina* and *junoša*), so that the latter are expected to be more error-prone than the former. In addition, the rate of acquisition for the rare male names in *-a* and the novel noun *obormoša* should be similar to those of low-frequency nouns *mužčina* and *junoša*.

Hypothesis III: The acquisition of the semantic principle proceeds in a rule-based fashion (cf. RC model, Yang 2002).

Prediction 3: There should be no differences in the accuracy rates between high-frequency nouns (*papa*, *deduška*, and *djadja*) and low-frequency nouns (*mužčina* and *junoša*): the latter should occur with high accuracy rates, since they belong to the same class with the high-frequency nouns *papa*, *deduška*, and *djadja*. High accuracy rates are also expected for the rare male names in *-a* and the novel noun, as they are in the same class too.

6.3 Results

6.3.1 Results of the main study

The results presented in this section were obtained in three production experiments, i.e. Experiments 1, 2 and 3 (see Sections 5.6, 5.7, and 5.8). In these tasks I elicited adjectival (color adjectives) and verbal agreement with five male kinship terms, ten male names in *-a*, and the novel noun *obormoša*.

The results were counted by hand. Tables 10.3 to 10.9 in Appendix II show numbers of agreement production for individual children. When counting children's responses, I excluded all unclear cases and counted every occurrence of an agreement target with or without a controller. For *papa*-type nouns and the novel nouns the agreement targets were mainly adjectives, but also some verbs and pronouns. For the novel noun they were mainly verbs, but also pronouns. Finally for the male names in *-a* they were only verbs. The likelihood of a carry-over effect was rather small in Experiment 2 on male names, since the test items were presented first and the fillers followed after that. Note however, that since there was no proper introduction in this experiment, most of the children made mistakes with regard to

¹Note that the rate of overgeneralization errors is unlikely to be very high, since, as (Pinker 1995:115) suggests, "...overregularization should be an exception, not the rule, representing an occasional breakdown of the system ...".

the first picture. Therefore, these responses were excluded. With regard to Experiment 3 on the novel noun, the test items did not follow immediately one after another: before the other test item was shown, the child had a break counting the animals together with Elmo. Finally, in Experiment 1 on *papa*-type nouns it was hard to control for carry-over effects. Recall that the test items, i.e. male kinship terms in *-a* followed after a neuter noun with an ending *-o* being stressed. Yet, adjectival agreement with them sounds identical to the one with feminine nouns (cf. *zelěn[ə]j[ə] peró* ‘green feather_(N)’ vs. *zelěn[ə]j[ə] mama* ‘green mommy_(F)’), unless the ending is stressed, e.g. *golubóje peró* ‘light blue feather_(N)’. Therefore, I made sure in the experiment that the color of each test item was different from the color of the preceding filler. When counting I paid attention to two factors: color and inflection of the agreement target. If the agreements as well as the color terms were identical (recall that I did not correct the children when they used the wrong color term) such forms were excluded. For example, if **zelěn[ə]j[ə] papa* ‘green daddy_(M)’ followed after *zelěn[ə]j[ə] peró* ‘green feather_(N)’ or *zelěnyj papa* ‘green daddy_(M)’ followed after **zelěnyj peró* ‘green feather_(N)’ it was excluded. Thus, I only included those agreement forms with the test nouns which differed from the forms used with the neuter fillers in either one parameter or the other or both.

Some examples of children’s responses are given in (3). The structures in (3-a) and in (3-c) contain two tokens each, i.e. two agreement targets, while the structure in (3-b) has just one. Occasionally, target-deviant and target-consistent forms could occur in the same utterance, as shown in (3-a).

- (3) a. *oj djadja upala u menja i perevernulsja* (Kolya
oops uncle/man_(M) fall_{PST.F} at I_{GEN} and flip over_{PST.M}
3;1)
‘Oops, my uncle/man fell and flipped over.’
Target structure: *djadja upal_{PST.M}*
- b. *snegovika stroila* (Lena 3;1)
snowman_{ACC} build_{PST.F}
‘(Jan’ka) was making a snowman.’
Target structure: *Jan’ka stroil_{PST.M}*
- c. *obormoša zalez krasnyj na derevo* (Petya 2;8)
obormoša_(M) climb_{PST.M} red_M on tree
‘Red obormoša climbed the tree.’

Table 6.1 shows the overall result of agreement production for the three noun groups: common nouns, proper names and a novel noun. It is clear that the accuracy rates of common nouns and proper names satisfy Brown’s (1973)

criterion of ‘use of grammatical structure in 90% of obligatory contexts’, since they are 92.0% and 96.0% respectively. The accuracy rate of the novel noun is 87.9% which is comparatively close. The error rates are quite low: 8.0%, 4.0%, and 12.1% with common, proper and novel nouns respectively.

Table 6.1: Overall agreement production with common, proper, and novel nouns (25 children, age 2;6-4;0)

noun type	N corr. (%)	N err. (%)	N tot. 100%
common nouns (<i>papa</i> , etc.)	682 (92.0)	59 (8.0)	741
male names in <i>-a</i> (<i>Trenya</i>)	217 (96.0)	9 (4.0)	226
novel noun (<i>obormoša</i>)	357 (87.9)	49 (12.1)	406

In order to answer the question related to input frequency I first divide the data by individual lexical items. Table 6.2 provides the overall distribution of agreement errors across the five common nouns individually. It is clear that there is no serious discrepancy in agreement production for different common nouns. On the whole the target-deviant forms occur at very low rates across the five common nouns, as we see in Table 6.2. This may suggest that children have complete gender mastery. However, the developmental data presented in Table 6.3 and Figure 6.1 reveal a hidden asymmetry in the error rates across individual nouns.

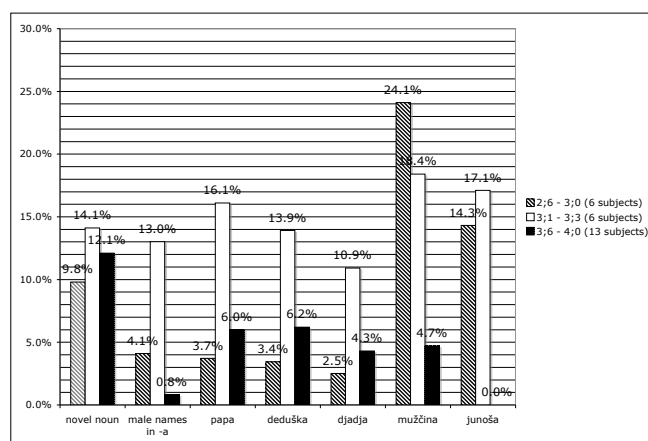
Table 6.2: Overall number and percentage of target-deviant agreement forms for five common nouns (25 children, age 2;6-4;0)

noun	N err. (%)	N tot. 100%
<i>papa</i> ‘daddy’	11 (7.7)	142
<i>deduška</i> ‘grandad’	10 (7.6)	130
<i>djadja</i> ‘uncle/man’	10 (5.5)	180
<i>mužčina</i> ‘man’	18 (11.7)	153
<i>junoša</i> ‘youth’	10 (7.3)	136

Table 6.3: Error rates of proper names, common and novel nouns expressed as number and percentage of the total production across three age groups (25 children, age 2;6-4;0). The first column includes number of errors/number of correct responses. The percentages of errors are given in parentheses

Age group	<i>obarmoša</i> N (%)	male names N (%)	<i>papa</i> N (%)	<i>deduška</i> N (%)	<i>djadja</i> N (%)	<i>mužčina</i> N (%)	<i>junosa</i> N (%)
2;6 - 3;0	8/74 (9.8)	2/47 (4.1)	1/26 (3.7)	1/28 (3.4)	1/39 (2.5)	7/22 (24.1)	4/24 (14.3)
3;1 - 3;3	13/79 (14.1)	6/40 (13.0)	5/26 (16.1)	5/31 (13.9)	5/41 (10.6)	7/31 (18.4)	6/29 (17.1)
3;6 - 4;0	28/204 (12.1)	1/130 (0.8)	5/79 (6.0)	4/61 (6.2)	4/90 (4.3)	4/82 (4.7)	0/73 (0)

Figure 6.1: Error rates of proper names, common and novel nouns expressed as a percentage of the total production across three age groups (25 children, age 2;6-4;0)

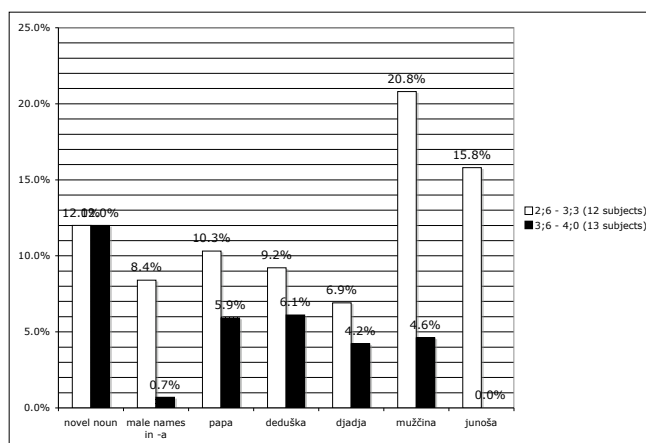


It is clear in Figure 6.1 that at the first age group (i.e. 2;6-3;0) the error rates for *mužčina* ‘man’ and *junoša* ‘youth’, are considerably higher than for other common nouns as well as for male names in *-a* (note that the latter are all considered together). The error rates of 24.1% and 14.3% do not allow me to conclude that their gender is fully mastered by younger children. For the next age group (i.e. 3;1-3;3) there is a sudden increase of errors across all nouns, but for *mužčina*. This noun exhibits lower accuracy rates than at the previous stage but they are still the highest compared to the rest of the nouns. Out of six subjects in the middle age group there are five who produced erroneous agreement forms with different nouns. Altogether 47 mistakes are found at this stage. It should be noted that the majority of the errors, i.e. 37, occur in the speech of one child (Kolya 3;1), who thus appears to be strikingly different from other children in the middle age group. Since children are known to “acquire language at widely varying rates” (Brown 1973:53), this could be an indication that the linguistic competence of this child does not correspond to the age of the other children in the middle age group, but may be more similar to the level of grammatical development of the younger children. I return to this finding in Section 6.4.1, where I suggest that age 3 might not be a reliable indicator of children’s grammatical development in the case of gender. Alternatively, it could be suggested that this child is simply more verbal than other children in his age group, i.e. he

produces more errors as well as more correct forms than other children. In fact, on the overall level almost half of his responses, i.e. 35, are correct. Yet, this evidence seems inconclusive to me, since many children in his age group are very productive and none of them makes as many errors as this child. Finally, with regard to the middle age group, we see that the asymmetry between *junoša* and *mužčina* and the rest of the nouns, which is very clear for the youngest children, is not noticeable at the age of 3;1-3;3, as the middle graphs in Figure 6.1 illustrate. Note also that the error rates for *junoša* and *mužčina* still persist at the highest level as compared to other nouns.

To present the asymmetrical error pattern more explicitly, in Figure 6.2 I arranged the the data to cover two age spans instead of three (as in Figure 6.1): 2;6-3;3 and 3;6-4;0.

Figure 6.2: Error rates of proper names, common and novel nouns expressed as a percentage of the total production for two age groups: 2;6-3;3 and 3;6-4;0 (25 children age 2;6-4;0)



It is clear in Figure 6.2 that between the ages of 2;6 and 3;3 agreement with the nouns *mužčina* and *junoša* is considerably more error prone than with other common nouns. The error rates for *mužčina* and *junoša* are also higher than for male names in -a and even for the novel noun.

Finally, both Figure 6.1 and Figure 6.2 demonstrate that after the age of 3;6, the asymmetry between the low-frequency common nouns (*mužčina* and *junoša*) and the other test items disappears. In particular, the error rates decrease considerably for all nouns, except the novel noun. Here the devel-

opmental pattern does not change much across the age groups; nevertheless, it reveals that masculine agreement is used correctly with the novel noun at a level of almost 90%. Note too that children in the last age group made no errors with the noun *junoša* ‘youth’. This is unlikely to be the result of sampling, since, as reported in Table 6.3, children produced 73 correct agreement forms with *junoša*, i.e. approximately as much as with other common nouns. This result may just be a coincidence.

6.3.2 Results of the pilot test

In this section I would like to compare the results discussed in the previous section to the results of the pilot test where the same and even more striking asymmetry has been found. Recall from Chapter 5 that the pilot test was carried out with 15 monolingual Russian children between the ages of 2;3 and 4;2, who were tested individually at day-care centre *Detstvo* in Ivanovo, Russia. The nouns used in all three pilot tests had the same grammatical properties as the nouns used in the main study. The test conditions were also identical to those in the main study. The differences were as follows. First, the noun *deduška* ‘grandad’ was not included in the pilot test, thus there were just four common nouns: *papa* ‘daddy’, *djadja* ‘uncle/man’, *junoša* ‘youth’, and *mužčina* ‘man’. Second, only adjectival agreement was elicited with the common nouns, while in the main study both verbal and adjectival forms were elicited with *papa*-type nouns. Third, in the pilot test I used familiar male names in *-a*, such as, for example, *Petya*. To avoid the familiarity effect very rare male names were used in the main study. It was thus unlikely that children could know them from before. Yet, as I show below, the children were equally good with both familiar and rare names. Finally, the novel noun *čunja* used in the pilot study was disyllabic and may have looked like a proper name. To make the noun more complex and similar to the low-frequency nouns *mužčina* and *junoša*, it was changed to *obormoša* in the main study. Thus, the changes made in the main study aimed to capture the familiarity effect and to balance the syllabic and sound structure of familiar vs. rare/novel test items.

In Table 6.4 I compare the overall results of the pilot test to overall results of the main study presented in Table 6.1 and repeated here for convenience. As we see there is little difference in the accuracy rates between the pilot test and the main study. Note, however, that children’s production for the novel noun *čunja* and common nouns in the pilot test is slightly worse.

Table 6.4: The accuracy rates for common, proper, and novel nouns in the pilot test (15 children, age 2;3-4;2) and the main study (25 children, age 2;6-4;0). Overall results

noun type	pilot test		main study	
	corr. N (%)	tot. 100%	corr. N (%)	tot. 100%
common nouns	165 (85.9)	192	682 (92.0)	741
male names in <i>-a</i>	115 (96.6)	119	217 (96.0)	226
novel noun	114 (78.6)	145	357 (87.9)	406

When the common nouns from the pilot test are considered individually, as shown in Table 6.5, we see that the error rates for *junoša* and *mužčina* are considerably higher than for *papa* and *djadja*. A similar asymmetry was found in the data from the main study, although it was only visible when the developmental data were considered, as shown in Figures 6.1 and 6.2.

Table 6.5: The error rates for individual common nouns in the pilot test (15 children, age 2;3-4;2)

noun	errorneous		total
	N	(%)	100%
<i>papa</i> ‘daddy’	1	(2.7)	37
<i>djadja</i> ‘uncle/man’	0	(0)	38
<i>junoša</i> ‘youth’	8	(14.3)	56
<i>mužčina</i> ‘man’	18	(29.5)	61

Importantly, the results of the pilot study in general replicate the findings in the main data set: they also reveal that children’s agreement production for *junoša* and *mužčina* is considerably worse than for *papa* and *djadja*. Another important fact is that both data sets reveal very high accuracy rates for male names in *-a* no matter whether they are familiar as in the pilot test, or rare, as in the main study, compare: 96.6% vs. 96.0% respectively.

6.3.3 Summary

The overall results show that children’s production is highly adult-like (roughly 90%) for all three noun groups, i.e. common, proper and novel nouns. However, the developmental data reveal a contrast between the individual items, which is most explicit at the age of 2;6-3;0 (cf. Figure 6.1). The errors for the

nouns *junoša* and *mužčina*, which are responsible for the asymmetry, persist till the age of 3;3. I suggested that this fact can be ascribed to the error production on the individual level, i.e. the level of children's grammatical knowledge can vary, so that some three-year-olds perform at a level similar to the two-year-olds. This may not be surprising, since it is well-known that some children are more advanced learners than others. Although the error rate for the novel noun slightly exceeds the allowed 10% error margin, as shown in Figures 6.1 and 6.2, it stays rather stable across the age groups. Finally, children of all ages show near adult-like knowledge of gender assignment for the rare male names in *-a*. It should be noted that the agreement production for rare proper names parallels the figures for *papa* 'daddy', *djadja* 'uncle/man', and *deduška* 'grandad', but differs from *junoša* 'youth' and *mužčina* 'man'.

6.4 The acquisition of gender through the prism of children's overregularization errors

In the following sections I discuss the experimental results presented in Section 6.3 with regard to the hypotheses and predictions formulated in Section 6.2. I first show that age 3 is not a reliable predictor of the acquisition process due to the differences between the high- vs. low-frequency nouns and between the individual children. I further consider the implications of the children's data for the WR model and the RC model. Children's overregularization patterns examined across different age groups and individual nouns reveal that the semantic procedure involves some item-based learning as predicted by the WR model. Thus, Yang's (2002) idea of a 'free ride' effect is not supported. Token frequency is argued to be the factor responsible for the asymmetry in the overgeneralization pattern across the individual common nouns, i.e. high- vs. low-frequency *papa*-type nouns. However, frequency alone cannot explain the course of acquisition of the semantic rule, as it appears to have no impact on the acquisition of gender with male names in *-a*. In order to explain this other asymmetry I explore the differences in the semantic representation of common nouns vs. proper names.

6.4.1 Age 3;0

Hypothesis I suggests that age 3;0 is the point which marks full mastery of the semantic principle, so that after 3;0 the error rates should be within the experimental error margin of 10%. One should keep in mind here that this assumption is based on the diary study of a single child, therefore the

data presented in this study may reveal a more general pattern. The results of the experimentation indicate that age 3;0 is not a good predictor of the time of acquisition of the semantic rule. First, the developmental evidence (Figure 6.1) shows that the adult-like (90%) capacity for the application of the semantic principle appears to be present in children long before 3;0. It is clear in Figure 6.1 that between the age of 2;6 and 3;0 not only high-frequency nouns (*papa* ‘daddy’, *djadja* ‘uncle/man’, and *deduška* ‘grandad’), but also rare proper names and a novel noun are overregularized at low rates, which do not exceed the experimental error margin of 10%. Hence, it can be argued that already at the age of 2;6, and presumably even earlier, Russian children have productive knowledge of the semantic rule which is generalizable to novel words. However, the agreement production at a rate of 75.9% and 85.7% for *mužčina* ‘man’ and *junoša* ‘youth’ respectively, cannot be considered highly adult-like during this early period. This means that before age 3;0 the semantic rule is established for some, but not for all relevant nouns. For the low-frequency common nouns it seems to be fully mastered only after this age. According to Figures 6.1 and 6.2 the acquisition of the semantic rule is complete for all nouns between the ages of 3;6-4;0, when the accuracy rates do not exceed the 90% level. These observations lead to the examination of the other two alternative hypotheses which address the role of input frequencies in gender acquisition and will be considered in the next section.

There is another reason to think that age 3;0 is not a reliable predictor of the developmental trajectory, since, due to individual differences between the children, overregularization errors can persist after this age for all the nouns. As pointed out in Section 6.3.1 with regard to Figure 6.1, one child (Kolya 3;1) seems to be responsible for the increase of the error rates across all noun types between the ages of 3;0 and 3;3. According to the individual results in Tables 10.3 to 10.9 in Appendix II, this child is very productive, but markedly worse than the other children in his age group. Clearly his poor performance is systematic and it is not restricted to one particular noun class; rather it spreads somewhat equally across all of them. At the same time this child is able to produce many target-consistent forms. For high-frequency nouns *papa* ‘daddy’, *djadja* ‘uncle/man’, and *deduška* ‘grandad’ the correlation of correct vs. erroneous forms in his speech is 17/10 respectively and 8/12 for the low-frequency *mužčina* ‘man’ and *junoša* ‘youth’, which fits the overall picture, i.e. low-frequency nouns are more error-prone. Given this, it is possible that this child is at the same level of grammatical development as the children in the first age group, i.e. 2;6-3;0. That is why the developmental trajectory becomes more explicit when the children are divided into two equal groups, as was previously shown in Figure 6.2.

In conclusion, the Russian developmental data provide support for a cross-

linguistic observation that the establishment of the semantic principle is a gradual process, which in this study reaches target-consistent level for all the nouns at around 3;6. It is clearly a result of development. However, as mentioned above, there are other factors, rather than age, that cause the particular developmental pattern. These factors explaining children's overregularization tendencies are considered in the following sections.

6.4.2 The (non-)effect of frequency on the acquisition of the semantic rule

In this section I would like to consider the implications of the data for the two theories of morphological acquisition discussed above: the WR model (Pinker 1999) and the RC model (Yang 2002). Recall that these theories explain children's overregularization tendencies in terms of frequency. Before I discuss my data against the specific predictions, a note regarding the differences in the experimental conditions is in order here. As mentioned in Section 6.3.1, the results were obtained in three production experiments, which were designed to elicit different agreement types: verbal (past tense) with male names in *-a*, and adjectival plus verbal with *papa*-type nouns as well as with the novel noun. On the overall level, male names in *-a* yielded the highest accuracy rate of 96.0% as compared to 92.0% for common nouns and 87.9% for the novel noun (cf. Table 6.1). Since very rare male names in *-a* were tested in the experiment, the result cannot be due to the familiarity effect (i.e. children could not know the gender of these nouns from before). Another possibility is that more successful acquisition for male names in *-a* is due to the difference in the experimental conditions, such that it was easier for children to construct correct gender agreement on the verb (as was the case of male names in *-a*) than on the adjective (in the case of *papa*-type nouns and the novel noun).

Recall also from Chapter 3 Section 3.3 that there are several factors that could make children's production of the adjectival forms more problematic compared to the verb forms. The first one refers to Gvozdev's (1961) observation that past tense verb forms appeared earlier (at approximately 1;10) than adjectival forms, which started being used productively at around 2;0. Nevertheless, he mentions that by the age of 2;4 his son had no difficulty with either adjectives or verbs in terms of gender. Another factor that could play a role here has been mentioned by Voeykova (1997). She has pointed out that adjectival endings are more diverse and complex as compared to verbal inflectional forms, which are simply $-\emptyset$ for masculine and *-a* for feminine (cf. Table 2.2 Chapter 2). This argument seems rather reasonable.

Nevertheless, there are at least two facts which suggest that the differences in the experimental conditions are unlikely to have an impact on the results in this study. First, it seems unlikely that at the age of 2;6 children can have problems with constructing adjectival agreement with color terms.² Second, an analysis of the data with respect to different agreement types reveals that past tense agreement is not necessarily less error-prone than adjectival agreement. In fact, as shown in Table 6.6, children perform considerably better with adjectives than with verbs for common nouns; the error rates are 4.8% for the adjectives vs. 17.3% for the verbs. With regard to the novel noun, verbal agreement is marginally worse than the adjectival, 11.5% vs. 14.2%, as shown in Table 6.7.³

Table 6.6: Agreement production with common nouns across different agreement types (25 children, age 2;6-4;0)

Agreement type	N corr. (%)	N err. (%)	Total (100%)
Adjective	514 (95.2)	26 (4.8)	540
Verb	129 (82.7)	27 (17.3)	156
Personal pronoun	39 (86.7)	6 (13.3)	45

Table 6.7: Agreement production with the novel noun across different agreement types (25 children, age 2;6-4;0)

Agreement type	N corr. (%)	N err. (%)	Total (100%)
Adjective	121 (85.8)	20 (14.2)	141
Verb	184 (88.5)	24 (11.5)	208
Personal pronoun	52 (91.2)	5 (8.8)	57

With regard to the proportions of errors reported in Table 6.6 the difference between adjectival and verbal agreement is statistically significant ($p=0.0985$, $p\leq 0.1$). This means that in the case of *papa*-type nouns, children perform significantly worse with verbs than with adjectives, i.e. contrary to the predictions mentioned above. Note, however, that the result

²According to Gvozdev (1961:438), color adjectives appear in his son's speech already at the age of 2;0 together with some other adjectives, although their total number is rather small; 23 adjectives are found in Zhenya's vocabulary before the age of 2;6. Note also that in Experiments 1 and 3 on *papa*-type nouns and the novel noun, I did not control for the correctness of the color terms used. Children could use any color they remembered. Very few of them (the youngest children) were operating with two or three, but many could use up to five or six.

³Pronominal agreement was not taken into account here, since the sample size is very small compared to the other agreement types.

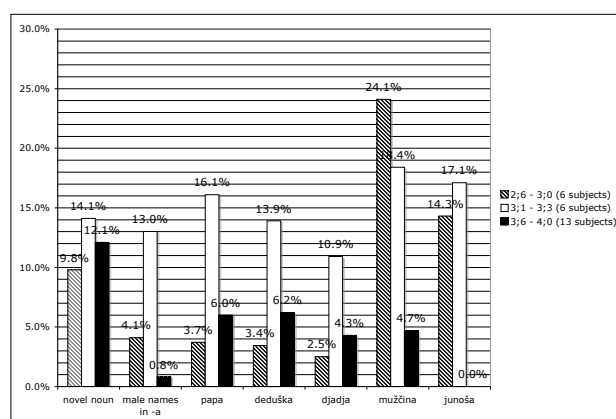
here may be due to sampling, since the number of structures produced with each agreement type differs considerably. In Table 6.7, on the other hand, where the samples of verbal vs. adjectival agreement are relatively similar in size, there is statistical evidence that the proportion of error rate is not different ($p=0.6836$, $p \leq 0.1$).

To conclude, children's better performance for rare male names in *-a* as compared to their performance for *papa*-type nouns and the novel noun, cannot be due to the difference in the experimental conditions, i.e. verbal agreement vs. adjectival. In Section 6.4.3 I propose an account of the asymmetries in children's agreement production for proper names vs. common nouns that bears on the differences in their semantic representation.

The role of frequency in gender acquisition

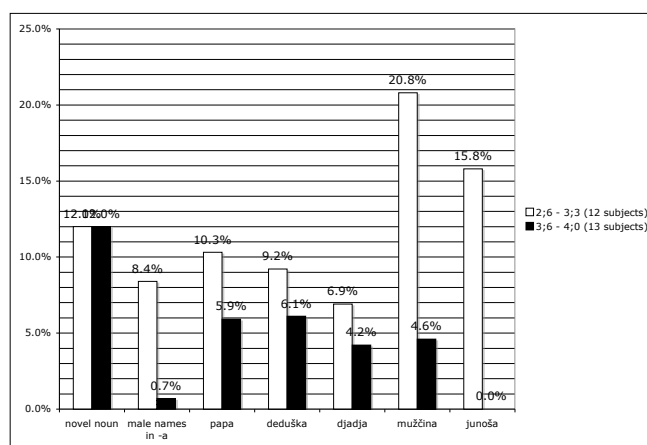
Hypothesis II, formulated along the lines of the WR model (Pinker 1999), predicted that children would show higher error rates for low-frequency nouns *mužčina* 'man' and *junoša* 'youth' than for high-frequency nouns *papa* 'daddy', *djadja* 'uncle/man', and *deduška* 'grandad'. In Figure 6.3 (repeated here from Section 6.3.1 for convenience) it is immediately evident that the first part of the prediction is borne out. Between the ages of 2;6 and 3;0 the percentage of errors for *mužčina*, i.e. 24.1%, and *junoša*, i.e. 14.2%, differs considerably from 3.7% for *papa*, 3.4% for *djadja* and 2.5% for *deduška*.

Figure 6.3: Error rates of proper names, common and novel nouns expressed as a percentage of the total production across three age groups (25 children, age 2;6-4;0)



This difference is not very pronounced at the next stage, i.e. between 3;1 and 3;3. However, if the data from the two earlier stages are collapsed, the contrast remains, as repeated in Figure 6.4: the percentage of errors for the high-frequency nouns is within the experimental error margin, i.e. 10%, while agreement production with the low-frequency nouns yields a higher percentage of target-deviant forms.

Figure 6.4: Error rates of proper names, common and novel nouns expressed as a percentage of the total production for two age groups: 2;6 - 3;3 and 3;6 - 4;0 (25 children, age 2;6-4;0)



In addition, a *t* test yields a significant result ($p=0.032$, $p \leq 0.1$), i.e. the error rates for high-frequency nouns *papa*, *djadja*, and *deduška* are significantly different from the error rates for low-frequency nouns *mužčina* and *junoša*. Such variability in the amount of target-deviant agreement forms, and such selectivity in which nouns are affected by errors, suggest that token frequency plays a role in the gender acquisition with these nouns. The error pattern observed in Figures 6.3 and 6.4 is compatible with the dual-process view on the acquisition process. That is, the children first learn how to use the semantic information for some frequent nouns and later extract the semantic rule, according to which sex-differentiable nouns denoting males are masculine. Finally, they generalize it to infrequent nouns in the same class.

What is left unexplained is the agreement production for the novel noun *obormoša* and male names in *-a*. With regard to Hypothesis II it was also predicted that the error rate for the rare male names and for the novel noun

should be higher than for the high-frequency nouns *papa*, *djadja*, and *deduška*. However, the evidence in Figures 6.3 and 6.4 does not support this prediction: children's production for male names in *-a* is highly adult-like (the error rate is 13.0% at its highest across the age groups), and surprisingly good for *obormoša* (the percentage of errors does not raise above 14.1% across the age groups); moreover it is unexpectedly better than for *mužčina* 'man' and *junoša* 'youth'. Thus, the evidence with regard to Hypothesis II is not straightforward and needs further discussion, which is provided in Section 6.4.3 below.

The alternative Hypothesis III formulated along the lines of the RC model (Yang 2002) predicted that the low-frequency nouns *mužčina* and *junoša* would occur with high accuracy rates, as they belong to the same class as the high-frequency nouns *papa*, *deduška*, and *djadja*. The same should be the case with the rare male names in *-a* and the novel noun *obormoša*, as they are in the same class too. In terms of evidence, only the second part of this prediction is borne out, which suggests that even at the early period, i.e. between 2;6 and 3;0, the child's mechanism for gender acquisition is not limited to initial rote learning. However, it is clear that the low-frequency nouns *mužčina* and *junoša* do not get a 'free ride' due to the high-frequency nouns attested in the same class. Thus, children's overregularization tendencies seem to be constrained by the nouns' individual frequencies in the sense of Pinker (1999), but not by the frequencies of a noun class in the sense of Yang (2002).

To conclude, the findings are contradictory and do not point towards a single solution. On the one hand, the contrast between high-frequency and low-frequency common nouns suggests that gender acquisition is a lexically-specific process and that children are sensitive to the frequency of exposure. On the other hand, the high accuracy rates for the novel noun and rare proper names indicate that rule-based acquisition is involved. The fact that rare proper names are overregularized at very low rates in contrast to the rare common nouns suggests that token frequency may not be the only factor that constrains children's overregularization tendencies. In what follows I explore other explanations of the pattern.

6.4.3 Proper names vs. common nouns in gender acquisition

In the previous section I concluded that nouns' proneness to overregularization can be partially attributed to token frequency. However, frequency seems to play a role with common nouns but not with proper names. Why

should this be so? My explanation is related to the fact that the semantic representation of proper names differs crucially from the semantic representation of common nouns. First, a proper name picks out a specific individual, while a common noun introduces a kind of individual (an individual of a class). Second, unlike a common noun, which has an indefinite number of referents, a proper name has just one (cf. Bloom 2000). For example, if a dog in the corner is *Fido*, then another dog (or any other animal) that walks in cannot be *Fido*, regardless of how similar they are. Of course, there can be several dogs called *Fido*, but *Fido* is still a proper name that refers to just one individual. In this case *Fidos* should be thought of as different words (Bloom 2000:126). Most importantly, a proper name appears to lack internal semantic structure, i.e. it does not describe the object it refers to, as a common noun does (cf. Burge 1973). For example, if someone is called *Vanya*, the hearer can infer that the person called *Vanya* is a male, while it is not simply a male that comes in association with *mužčina*, but an adult male. Thus proper names form a discrete semantic class, distinct from the class of common nouns.

In the acquisition literature it has been noticed that proper names have a special status in child grammar (cf. Macnamara 1982, Gentner 1982, Bloom 1990; 2000): not only do they appear among the very first words of children learning different languages (such as German, English, Turkish, Japanese, Mandarin, etc. (cf. Gentner 1982)), “children learn their first proper names for people long before they learn any common noun that refers to these individuals (such as *person* or *parent*)” (Bloom 2000:130). Various experimental studies reveal children’s early understanding of the syntax and semantics of proper names. Macnamara (1982) was the first to show experimentally that seventeen-month-old English-speaking girls are sensitive to the syntactic cues in structures like “This is *zav*” and “This is a/the *zav*”.⁴ The former is interpreted by them as a proper name, while the latter as a name for the kind. Interestingly, they notice the presence/absence of the articles before a name only when the object is a doll, not when it is a block. In other words, the syntax is irrelevant when *zav* is used to refer to an inanimate object. Apart from syntax, children seem to be aware that an object gets only one proper name (Hall and Graham 1997) and that if a word refers to more than one object, it is unlikely to be a proper name (Hall 1996) (rather than a common noun). Thus, from early on, even before they utter their first words, children seem to be familiar with the concept of a proper name, such that it picks out a specific individual (human beings, as well as animals or toys), and to be

⁴Boys reveal sensitivity to syntax at the age of 27 months. Macnamara ascribes this delay to the boys paying less attention during the test than the girls.

able to distinguish it from a common noun, which has an indefinite number of possible referents.

Finally, children also seem to know that proper nouns name objects but do not describe them. Naming specific entities is shown to be a characteristic feature of children's early vocabulary acquisition (Gentner 1982). In the speech of one American English-speaking child, Gentner observes that the set of his first words includes the names of individual entities, where even common nouns appear to be names of particular entities, e.g. *dog* is used to refer to a particular dog and *duck* to a small ceramic object (that was in fact a chicken). At the beginning, *Daddy* and *Mamma* are also used by this child for appropriate individuals, however very soon they are generalized to other similar-appearing men and women. To conclude, even common nouns are first used by children to refer to specific objects or individuals, just like proper names, i.e. with no intention to describe them.

It can thus be predicted that the semantic function of proper names and their special status in child language are the factors that might facilitate gender acquisition with these nouns. As we know, this prediction is borne out by the fact that the error rates for the rare proper names are considerably lower compared to the rates for the rare common nouns. I would like to propose the following explanation of the phenomenon. Consider first that in order to assign masculine gender to proper and common nouns a child should establish their semantic content. Although all of them denote a male, *papa*-type nouns have additional meaning, while proper names do not. (*Papa*, *deduška*, and sometimes *djadja* are kinship terms, which imply a kind of family relation. *Mužčina* represents an adult male, and *junoša* a young male.) This additional semantic content cannot be learnt on a single exposure, therefore frequency comes into play. Proper names, on the other hand, lack this ability to describe, therefore they are usually learnt on a single exposure. As I have argued above, the concept of a proper name is familiar to the children from the outset. This knowledge, I suggest, helps more than frequency in gender acquisition of these nouns. In other words, frequency does not play a role in the case of proper names, where the lack of additional semantic information ensures successful application of the semantic principle in gender assignment by children. With regard to common nouns, frequency and meaning seem to be interrelated. Recall, for example, that gender agreement with, *mužčina* was more error-prone than with *djadja*. Both of them denote adult males, i.e. they have the same meaning but differ in frequency of occurrence.⁵ *Djadja* is frequent, therefore its semantics is learnt first, and so is the gender. *Mužčina*

⁵ *Djadja* can also mean 'uncle', but in child directed speech it is generally used to denote a 'man'.

is rare, therefore its mapping takes longer, and so does the gender.

Another possible factor that may have caused a difference between the rare proper names and the low-frequency common nouns *mužčina* and *junoša*, could be children's sensitivity to certain combinations or 'frames' in the input, such as 'common noun + proper name'. It is a characteristic feature of Russian child-directed speech to use a proper name with nouns like *deduška* 'grandad' and *djadja* 'uncle/man' in order to distinguish between different individuals, e.g. *deduška Vasya* or *djadja Kolya*. Although this pragmatic reasoning does not apply to *papa* 'daddy', this noun is often used in combination with a proper name, e.g. *papa Miša*, in kindergartens when the teachers distinguish between the fathers of different children.⁶ I suggest that the (frequent) occurrence of the nouns *papa*, *deduška*, *djadja* in combination with a proper name in the input may be responsible for children's more accurate agreement production with these nouns as opposed to the agreement production with the nouns *mužčina* 'man' and *junoša* 'youth', which are never used in combination with a proper name. Comparing the error rates of the nouns *djadja* and *mužčina*, which share the same content, i.e. 'man', in Figures 6.3 and 6.4, reveals that between 2;6 and 3;3 the former is considerably less error-prone than the latter. This may be due to the fact that the former but not the latter is often used in combination 'common noun + proper name' by adults.

In conclusion, token frequency cannot explain children overregularization pattern with the rare proper names vs. rare common nouns. Therefore, I proposed that children's early awareness of a nouns' semantic representation and hence their ability to make a distinction between common vs. proper nouns constrains children's overregularization tendencies. Finally, the occurrence of certain common nouns with proper names in the input may be an additional factor. However, this idea needs further empirical support from future research.

⁶In the sample of the mother's data from Varvara's corpus discussed in Chapter 4 Section 4.2.3, I found three occurrences of the noun *deda* 'grandad' with a proper name, e.g. *a čto tebe deda Saša podaril?* 'And what did grandad Saša give to you?' However in this corpus there were no relevant examples of *papa* and *djadja* used with a proper name.

A Google search for "papa Saša" reveals 3380 examples and for "djadja Saša" 82900. 'grandad + proper name' appears to be less frequent in Google: 585 for "deduška Saša" and 623 for "deda Saša". For "mužčina Saša" and "junoša Saša" it reveals 311 and 214 examples respectively, although the majority of those appear to be irrelevant on the close look, since e.g. the two words occur on the clause boundaries, i.e. they do not constitute a noun phrase.

6.4.4 Blocking as an explanation for the overregularization and the course of gender acquisition

In the previous sections I have shown that children gradually establish the target gender representations for *papa*-type nouns and male names in *-a*. The transition from formal to semantic assumed at the beginning of the chapter becomes evident from the children's overregularization patterns. Within the considered developmental period, 2;6-4;0, the gradual transition is most evident for the low-frequency nouns, with regard to which the error rates are declining gradually and slowly, so that children's production reaches the target-consistent level for all nouns at approximately the age of 3;6. Errors due to overregularization indicate a strong impact of morphology, which appears to be more dominant in younger than in older children. Thus the results discussed in this chapter provide support for the claim that children base their early hypotheses on the morphological properties of nouns and not on their semantic properties related to natural gender. The remaining question, which I consider in this section, is how and why children eliminate the wrong hypotheses and arrive at an adult-like gender representation where the semantic rule dominates over the morphological rule. The puzzle we are looking at is rather complex, since the developmental sequence does not simply progress from overregularization to correct performance. What we find is the simultaneous use of correct and incorrect agreement forms over a long period of time, which indicates that two alternative hypotheses are in competition. This is the essential problem of language acquisition, when a child has to realize that one form is ungrammatical in the absence of such information in adult speech. Input (or positive evidence) is known to contain limited evidence concerning the correct rules of the language. Hence it does not provide overt information about which forms are ungrammatical. Such information can be provided by parental feedback, also called (direct) negative evidence, in the form of adults' corrections, disapproval, etc. However, it has been shown that negative evidence has no significant effect on children's linguistic behavior, due to its weakness and inconsistency (Brown and Hanlon 1970, Demetras 1986, Marcus 1993). Without negative evidence, any model of language acquisition is challenged to explain how the child can learn language (more specifically, to unlearn grammatical errors) from positive evidence alone.

Since Chomsky (1959) the limited nature of the input (the poverty of stimulus) is often used as an argument in favor of a learning mechanism which is part of the child's innate universal grammar. For example, the blocking principle, also known as the 'elsewhere condition' (Kiparsky 1973), has been proposed by Marcus (1992) and Pinker (1995) among others as

a means of recovering from overregularization errors in the acquisition of the English past tense system. Blocking is the operation that governs the relations of two competing forms by forcing the use of a more specific form over a more general form in the absence of negative evidence. For example, an irregular form *held* is used, since it is a more specific realization of *hold* than a regular form *holded*. Blocking is thus claimed to be part of an innate universal grammar and a cause (but not the effect) of the child's language learning.

I propose that the idea of blocking could be used to account for how Russian children recover from the overregularization of feminine gender with *papa*-type nouns and male names in *-a*. However, if blocking is to be used, the gender assignment rules formulated by Corbett (1991) have to be readjusted. More specifically, if children initially make use of the morphological rule in (4-b) and assign feminine gender to all nouns in declension II they should later assume that not all nouns in declension II are feminine, i.e. there is a subset of nouns that denote males and hence must be masculine. This means that with regard to the subtypes of nouns considered in this chapter the semantic rule in (4-a) has to be reformulated, so that it defines a subset of forms generated by a more general morphological rule in (4-b) (the rules in (4) are repeated here from Section 6.2):

- (4) a. nouns denoting males are masculine;
 b. nouns of declensional class II are feminine.

A more specific variant of the semantic rule for these nouns might be as follows:

- (5) nouns of declensional class II referring to males are masculine

In terms of the blocking theory, the acquisition of gender with *papa*-type nouns and male names in *-a* involves a process of competition between the morphological rule in (4-b) and the semantic rule in (5). In order for these nouns to be assigned masculine gender correctly a more general morphological rule has to be blocked by a more specific semantic rule. Thus, lacking negative evidence, children can make use of the internal mechanisms like e.g. blocking to unlearn ungrammatical utterances.

Importantly, the postulation of a blocking principle for the acquisition of gender in Russian means that the assignment rules formulated by Corbett should be readjusted. With regard to the developmental sequence proposed above, more specific semantic rules are needed so that children can overcome errors and arrive at the adult state. In the course of my investigation the idea of specific semantic rules will receive considerable support from chil-

dren's agreement production with other subcategories of nouns. In Chapter 9, based on these findings, I will propose a cue-based approach to gender acquisition. In particular, pieces of structure in child's internal grammar or cues (in the sense of Lightfoot 1999) will be argued to trigger the progression from overregularization to correct performance. Their formal representation will bear resemblance to the semantic rule I arrived at in (5). Most importantly, I will argue that there are separate semantic cues for the individual classes of nouns and that qualitative and quantitative properties of the input play an important role in their retrieval.

6.5 Summary and conclusion

In this chapter I have made an attempt at providing an explanation of how gender acquisition in conflict situations is actually achieved. The main concern of this chapter has been the integration of the semantic principle into the gender assignment process in obligatory contexts. The following main points arise from the examination of the acquisition process for two subclasses of nouns, i.e. male kinship terms in *-a* and male names in *-a*, as well as for the individual nouns. I have concluded that although the semantic rule is available to children before age 3;0, it appears to be fully mastered around 3;6. Furthermore, evidence from children's overgeneralization errors for individual *papa*-type nouns suggests that gender is sensitive to frequency, i.e. the course of acquisition of the semantic principle is constrained by the frequency of exposure. This result is thus consistent with the claim made by the Words and Rules theory that the acquisition of exceptional morphology proceeds from rote to rule, since the acquisition of the semantic principle is clearly the result of development. At the same time, the absence of the free-ride effect for the low-frequency nouns *mužčina* 'man' and *junoša* 'youth' in my study disconfirms the idea of the Rules and Competition Model that this should be a single, rule-based, process. Yet, I have proposed that there are additional constraints that can explain why the gender of some nouns, i.e. male names in *-a* and high-frequency male kinship terms in *-a*, is easier to acquire than the gender of other nouns in the same class. While token frequency has been claimed to be responsible for the gradualism in the acquisition of the semantic rule for *papa*-type nouns, another factor, such as the lack of internal semantic structure, seems to have a beneficial role in gender acquisition of proper names. Thus, children appear to be sensitive to the differences in the semantic concept of natural gender, whose representation has been established to be more complex in the case of common nouns than in the case of proper names. This result suggests that proper names form a

discrete semantic class distinct from other nouns. This is parallel to Roeper's (2007) idea that children function with great semantic subtlety and are sensitive to classes of categories, which has been discussed in Chapter 4 Section 4.5. Based on this, it has been predicted that the course of acquisition for different noun classes may be different, i.e. if children pay attention to the notion of class and generalize within it, different overgeneralization patterns are to be expected. These assumptions will be tested further in the following chapters where I consider the acquisition of hybrids and double gender nouns and compare it to the acquisition of nouns in obligatory contexts.

Finally, I proposed that children may be sensitive to input strings, such as 'common noun + proper name'. This factor has been suggested to account for higher accuracy rates with male names in *-a* and some male kinship terms in *-a*, i.e. *papa*, *deduška*, and *djadja* vs. *mužčina* and *junoša*. Yet, more evidence is needed to explore this assumption.

To summarize, based on the results of the novel data as well the previous acquisition research, I can conclude that the two gender principles - semantic and morphological - are qualitatively different. The summary of the main points is displayed in Table 6.8, where the morphological rule, despite being language specific, has the properties of the implicit rule, since it is acquired instantaneously and frequency appears to have no effect here. The semantic rule, on another hand, being based on the universal notion of natural gender, is developed slowly depending on frequency. Thus it appears to be experience-dependent.

Table 6.8: Comparison of gender assignment principles

Morphological criterion	Semantic criterion
language specific	universal male/female distinction
semantically abstract	based on the semantic concept
acquisition is rapid/instantaneous	acquisition is gradual
depends of consistency, not frequency	depends on frequency
single process 'rule-based acquisition'	dual process 'from rote to rule'

After all the difference in the acquisition of the two criteria may be not so striking if one views morphology as a grammar internal component and semantics, i.e. the sex-based distinctions, as an extra-linguistic factor. Interestingly, as I show in the next chapter, the status of the semantic criterion, i.e. whether natural gender represents a grammatical vs. socio-cultural phenomenon, can also have an effect on the course of acquisition. In particular, when the application of the semantic rule is caused by socio-cultural factors its acquisition is delayed as compared to the cases where the semantic rule

is categorical.

Chapter 7

Acquiring gender in variable contexts

7.1 Introduction

In the previous chapter, focusing on *papa*-type nouns and male names in *-a*, I have shown how the acquisition of the semantic rule proceeds in obligatory contexts and what factors play a role in this process. In such contexts, where the semantic rule has categorical status, it appears to be used near-categorically already in the speech of two-year-olds. Nevertheless, we could see that the semantic rule takes over gradually, as for low-frequency nouns it was not fully mastered until approximately the age of 3;6.

The aim of this chapter is to examine the acquisition of the semantic rule in variable contexts where it does not have categorical status. On the basis of experimental evidence, I first show that the acquisition of the semantic principle is delayed in variable contexts compared to obligatory contexts. I argue that the delay is caused by the children's lack of socio-cultural constraints on use. I further compare the agreement patterns of two- and three-year-olds, their primary caregivers, and five- and six-year-olds, and show that children's socio-linguistic competence matures slowly.

Crucially, the results discussed in this chapter provide evidence for one of the major claims that I want to put forward in this dissertation, namely that children distinguish between classes of nouns. Detailed discussion of this issue will be postponed till Chapter 9, where I compare the course of acquisition for all subcategories of nouns.

This chapter targets the acquisition of two noun types: hybrids referring to females (e.g. *vrač* 'physician') and female names that take masculine suffixes *-ok* or *-ik* (e.g. *Verok*). Both noun types have ambiguous gender cues:

their morphology points towards masculine gender, while their semantics suggests feminine (i.e. exactly the opposite of *papa*-type nouns). In addition, unlike *papa*-type nouns, agreement, being variable, does not resolve the ambiguity here, since both masculine and feminine agreement forms are attested for these nouns in the adult language in reference to a female, as shown in (1). Thus children have to deal with input ambiguity (in terms of conflicting gender cues) and input variability (in terms of optional agreement forms).

- (1) a. *Vrač prišl-a* / *Vrač prišel-Ø*.
 physician come_{PST.F} / physician come_{PST.M}
 ‘The physician came.’
- b. *Natusik zabolet-a* / *Natusik zabolet-Ø*.
 Natusik get sick_{PST.F} / Natusik get sick_{PST.M}
 ‘Natusik got sick.’

For the examples in (1) the variability involves variable use of semantic and morphological principles and is realized via feminine and masculine agreement forms. As discussed in Sections 2.4.3 and 2.4.4, the choice of agreement with hybrids and female names in *-ok/-ik* is constrained by the type of agreement target and, as pointed out by (Corbett 1991:183), the likelihood of semantic agreement is high in verbal predicates. Therefore, in Experiments 4 and 6 (see Chapter 5 Sections 5.9 and 5.11), the results of which are considered in this chapter, I elicited verbal predicate agreement with these nouns.

It should be noted that there is one important difference between hybrids and female names in *-ok/-ik*. The latter denote females only, while the former can refer to either females or males. Therefore semantic feminine agreement with a hybrid, e.g. *Vrač prišl-a* as in (1-a), helps to avoid the ambiguity in this case (since feminine agreement is ungrammatical with a male referent), however, masculine agreement (*Vrač prišel-Ø*) can still imply not only a female referent, but also a male referent.

The situation where the semantic principle has categorical status for some nouns (e.g. *papa*-type nouns and male names in *-a*), and variable status for other nouns (e.g. hybrids and female names in *-ok/-ik*) gives rise to the question whether the semantic principle is acquired simultaneously in obligatory vs. variable contexts. Predictions related to this question were previously discussed in Chapter 4 and were based on certain theoretical assumptions as well as previous research on acquisition of variation in the input. In Section 7.2 of this chapter I formulate the hypotheses and predictions related to this question more specifically. In the following sections I evaluate these predictions with children’s and adults’ production data, and demonstrate that

children seem to initially construct a grammar that is different from their caregivers' grammar and where the semantic principle is not dominant. In Section 7.3 an overview of the findings in the child data is provided. In Section 7.4 I discuss children's agreement behavior in variable contexts compared to the acquisition process in obligatory contexts, discussed in the previous chapter. Adult and child agreement preferences are analyzed in Section 7.5. Finally, Section 7.6 focuses on the developmental issues. Specifically, I compare the status of the semantic principle in variable contexts in the speech of two- and three-year-olds and in the speech of five- and six-year-olds and conclude that older children are more aware of the socio-cultural impact of the semantic principle with these nouns. However, the course of acquisition is different for hybrids vs. female names due to some additional factors. The chapter ends with a summary of the conclusions in Section 7.7.

7.2 Hypotheses and predictions

In this section I formulate two alternative hypotheses for the acquisition of the semantic rule with hybrids referring to females and female names in *-ok/-ik*.

The first hypothesis follows from an observation that socio-cultural constraints on use are acquired later than grammatical constraints (cf. the discussion in Chapter 3 Section 3.4 and Chapter 4 Section 4.3).

Hypothesis I: Given that the use of the semantic rule for hybrids referring to females and female names in *-ok/-ik* is motivated by socio-cultural factors, its acquisition will be delayed.

Prediction 1a: The course of acquisition of the semantic rule in variable vs. obligatory contexts is expected to be different: in variable contexts children are expected to use the semantic rule considerably less often than in obligatory contexts.

Prediction 1b: Adults' and children's production patterns are expected to be different: the semantic rule will be used predominantly by adults, but infrequently by children.

In terms of agreement, both Prediction 1a and Prediction 1b suggest that semantic feminine agreement should occur rarely, while syntactic masculine may prevail.

The alternative hypothesis is as follows:

Hypothesis II: The semantic principle is acquired simultaneously for all nouns: as soon as children establish the dominance of the semantic principle for one noun class, in other words, as soon as they realize the grammatical

function of the semantic principle for one subcategory of nouns (e.g. *papa*-type nouns) they must be able to extend this knowledge to other noun classes.

Prediction 2: The semantic principle for nouns in variable contexts should be acquired at the same time as for nouns in obligatory contexts.

7.3 Results

The results presented in this section were obtained in Experiments 4 and 6. Experiment 4 was designed to elicit verbal predicate agreement with female names in *-ok*, *-ik* and Experiment 6 aimed at eliciting noun-verb agreement with hybrids referring to females. Recall that hybrids whose referents were males were included in the test as fillers. The same experiments were conducted with the children's primary caregivers (parents, grandparents or older siblings), as well as 12 older children aged 5;1 - 6;5. In what follows I first present the results from the children in the main study. I then consider the results from caregivers and older children.

Tables 10.10 and 10.13 in Appendix II present the individual results of agreement production for 25 children aged 2;6-4;0. When counting the children's responses, I excluded all unclear cases and counted every occurrence of an agreement target with or without a controller. Some examples with overt and covert controllers are given in (2) and (3).

- (2) a. lenok narjadilsja (Seva 3;0)
 Lenok dress_{PST.refl.M} up
 'Lenok dressed up.'
 b. platje nadela (Roma 2;10)
 dress put_{PST.F} on
 '(Lenok) put on a dress.'
- (3) a. doktor polečil (Kolya 3;1)
 doctor cure_{PST.M}
 'Doctor cured him.'
 b. mjačik dala (Vova 3;7)
 ball give_{PST.F}
 '(The policewoman) gave the ball (to him).'

Tables 7.1 and 7.2 provide an overview of the children's utterances with female names and hybrids respectively where the subject element is either present (NV) or not (₋V), as well as those where the subject is a third person pronoun (proV). The structures containing a pronoun and a verb were very few and were excluded from further analysis. This was done in order to have

a more clear picture with regard to the noun-verb agreement. As I show below, the verb agreeing with the pronoun always has a matching form, e.g. *ona dala* ‘she give *PST.F*’. This means that when the child uses pronoun *ona* ‘she’ which has a clear feminine form (while the test noun it refers to has masculine) there is a possibility that gender agreement on verb corresponds to the form of the pronoun rather than to the sex of the referent. In other words, it may be easier for a child to produce feminine agreement on the verb, because of the pronoun *ona* ‘she’, whose morphological form suggests feminine, hence there is no need to consider the sex of the referent of the test noun. Given these observations, I excluded structures containing both *ona* ‘she’ as well as *on* ‘he’ to have more clear results.

	M	F	Total
NV	34	8	42
	81.0%	19.0%	100%
_V	157	24	181
	86.7%	13.3%	100%
proV	6 (he)	1 (she)	7

Table 7.1: Children’s use of NV, _V, and proV structures with female names in *-ok*, *-ik* (25 children, age 2;6-4;0). Number and percentage of tokens

	M	F	Total
NV	7	1	8
	87.5%	12.5%	100%
_V	239	56	295
	81.0%	19.0%	100%
proV	3 (he)	3 (she)	6

Table 7.2: Children’s use of NV, _V, and proV structures with hybrids referring to females (25 children, age 2;6-4;0). Number and percentage of tokens

It is clear that the structures where the subject is covert (_V) are in majority: 181 with female names and 295 with hybrids (cf. NV 42 and 8, and proV 7 and 6 respectively). The structures with overt subjects are less common: 42 for female names and 8 for hybrids. Children’s tendency to omit the test nouns in the experimental situation is not surprising, since in Russian, which is often considered an optionally null subject language (Gordishevsky and Avrutin 2004) or a discourse-oriented pro-drop language (Zdorenko 2005), the omission of elements (subjects or direct objects) is natural when their referents are easily recoverable from the context. According to Gordishevsky and Avrutin (2004) among other researchers, the omission of subjects in pragmatically motivated contexts is a characteristic feature of Russian child and adult language. Note that in the experimental context provided to the children in my study, the test nouns (and their referents) were always available to the participants. Sometimes, when children had difficulty accessing the target structure, the experimenter could provide the beginning of it (e.g.

I Lenok... ‘And Lenok...’), where the test noun was instantiated. In this case the noun appeared in the immediate context, which is most suitable for subject omission. An interesting question which arises with regard to the structures with omitted elements is whether these are nouns or pronouns.¹ Although a detailed investigation of this issue is beyond the scope of this thesis, since my main interest is in how agreement corresponds to the sex of a referent, I would like to make the following observation. Given Corbett’s (1991) Agreement Hierarchy (see (17) in Chapter 2), semantic (feminine) agreement is predicted to be used on personal pronouns. In other words the pronoun *ona* ‘she’ should be used with hybrids referring to females and female names in *-ok/-ik*. This was in fact found in the caregivers’ production, who only used *ona* ‘she’, which corresponded to the sex of a referent (altogether caregivers produced three proV utterances with female names and 11 with hybrids). If *ona* ‘she’ is the target-consistent alternative, then masculine agreement in the structures with covert subjects (i.e. $_V$) in Tables 7.1 and 7.2 may indicate that children in fact omit nouns, since they have masculine shape. Yet, it is also possible that children omit the pronoun *on* ‘he’, which also has a masculine shape, since it ends in a hard consonant. This can be true, since, unlike caregivers, the children in my study were found to use both *ona* ‘she’ and *on* ‘he’ for females, as illustrated in (4-a) and (4-b). Note that in (4-b) the child produced a sequence of noun-pronoun-verb with masculine agreement on both targets, which does not correspond to the sex of the referent. Moreover, the example in (4-c) demonstrates that syntactically agreement between the pronoun and the verb was always correct; however children could hesitate in the choice of a pronoun. This example suggests that the first choice, i.e. *on* ‘he’, which shows the influence of the noun’s form, is immediately rejected by the child in favor of a the semantically appropriate variant, i.e. *ona* ‘she’. This was a single occurrence in the speech of one child (Polya 3;10) who could use both masculine and feminine verb forms (4/6) for female names but only masculine for hybrids referring to females (13/0).

- (4) a. *ona šariki mame prinesla* (Lyuba 3;3)
 she balloons to mommy bring_{PST.F}
 ‘She (Verok) brought the balloons for the mommy.’
- b. *lenok # on nadel šapku i vot ešče* (Slava 3;6)
 lenok he put_{PST.M} hat and this also
 ‘Lenok put on a hat and this as well.’

¹Following Rizzi (1982), it is generally assumed that null subjects in the languages like Italian are null pronouns, which are identified by the verbal inflection. Such null elements are called *pro*.

- c. on dal +// ona emu dala paločki (Polya 3;10)
 he give_{PST.M} she him give_{PST.F} sticks
 ‘S/he (a female doctor) gave her sticks.’

Note, in Chapter 8 I will show that unlike female names and hybrids, the children’s choice of third person pronouns with double gender nouns is always semantically motivated, i.e. they only use *ona* ‘she’ when a referent of a double gender noun is a female and *on* ‘he’ only when it is a male.

Finally, with regard to Tables 7.1 and 7.2, it is most striking that children produce masculine agreement with female names in *-ok/-ik* and hybrids referring to females more than 80% of the time both when the subject is overt (i.e. NV) and when it is covert (i.e. *_V*). Clearly, semantic (feminine) agreement is a less preferred option in both types of structures. Interestingly, on the individual level, some children could use both masculine and feminine agreement when the same subject was missing, as shown in (5). Two children did so in the task on female names and the same two children plus two other in the task on hybrids.² These children produced one masculine and one feminine verb form, as illustrated in (5), except one child (Nadya, age 3;3) who produced one feminine and three masculine forms in reference to a female doctor.

- (5) a. zakryl glazki (Nadya 3;3)
 close_{PST.M} eyes
 ‘(Verok) closed her eyes.’
 b. razbudilas’ (Nadya 3;3)
 awake_{PST.refl.F}
 ‘(Verok) woke up.’

Table 7.3 displays the overall agreement production with female names in *-ok*, *-ik* and hybrid nouns referring to females from 25 children that participated in the experimentation (as mentioned above, the structures consisting of a personal pronoun and a verb are excluded here). The bottom line in Table 7.3 illustrates agreement production with hybrids referring to males; these data have been obtained for control.

²These children are Kolya 3;1, Nadya 3;3, Lyuba 3;3, Roma 2;10.

Table 7.3: Overall agreement production with female names in *-ok/-ik* and hybrid nouns referring to females (25 children, age 2;6-4;0)

	M N (%)	F N (%)	Total (100%)
Female names in <i>-ok/-ik</i>	191 (85.7)	32 (14.3)	223
Hybrids referring to females	246 (81.2)	57 (18.8)	303
Hybrids referring to males	251 corr. (96.9)	8 err. (3.1)	259

A number of main points emerge from Table 7.3. First, on the overall level, children's production with both hybrids referring to females and female names seems to be variable, since the children can use syntactic (masculine) agreement as well as semantic (feminine) agreement. At the same time the examination of the individual data reveals that not all children use variable forms (see Tables 10.10 and 10.13 in Appendix II). More specifically, only nine children (out of 25) demonstrate variation between masculine and feminine for female names³ and 18 for hybrids. The other children, i.e. 15 in the case of female names³ and seven in the case of hybrids, use only one form, which is masculine. This particular finding suggests that in the case of female names children do not favor variability, which, as I conclude below, is true about their caregivers as well. Interestingly, however, caregivers' agreement preferences are quite different: unlike children they use feminine near-categorically with female names and predominantly with hybrids referring to females. A more detailed comparison of child and caregiver production is provided in Section 7.5.

Second, it is clear in Table 7.3 that children are not prone to employ the semantic rule in gender assignment with hybrids referring to females and female names in *-ok/-ik*, since they produce semantic (feminine) agreement with them rather infrequently: at a rate of 14.3% for hybrids and 18.8% for female names. On the contrary, children appear to be prone to use the morphological rule for both subcategories of nouns, since masculine agreement forms constitute 81.2% and 85.7% of the total number of utterances produced for hybrids referring to females and female names respectively. Thus, there is a tendency among the children to produce masculine agreement with these nouns, even when they are shown a picture of a female. Clearly, children's bias towards masculine with hybrids referring to females and female names emerges from their sensitivity to the noun's morphological form, which has been established by previous acquisition research on gender in Russian as

³Note that the data from the youngest child (age 2;6) contains only one agreement form for female names.

well as in various other languages (cf. the discussion in Chapter 3). The fact that children focus more on the shape of the nouns than on their semantics in variable contexts is the most important characteristic of the children's production, which, as I show below, makes it considerably different from the adult language.

Finally, as expected, with hybrids referring to males children use correct masculine agreement at a rate of 96.9%. Notice also that the number of errors for these nouns, 3.1%, is within the experimental error margin (i.e.10%) and these occur in the speech of three children only. Thus, it can be concluded that children's production in this case is not variable, but categorical, i.e. target-like.

To examine children's agreement preferences against the adults', caregivers' data have been analyzed in Table 7.4. Tables 10.11 and 10.14 in Appendix II present individual results for 21 adult speakers.⁴

Table 7.4: Caregivers' overall agreement production the female names in *-ok/-ik* and hybrids referring to females (21 adult speakers)

	M	F	Total
	N (%)	N (%)	(100%)
Female names in <i>-ok/-ik</i>	4 (2.2)	175 (97.8)	179
Hybrids referring to females	53 (21.5)	193 (78.5)	246

In Table 7.4 it is immediately clear that the caregivers construct agreement with female names and hybrids based on the semantic principle at a rate of 97.8% and 78.5% respectively. Thus, they are much more willing to employ the semantic principle for these nouns than their children, whose data presented in Table 7.3 reveals the reverse pattern of use.

The speech of the caregiver is considered to be the main source from which a child learns a language (Gallaway and Richards 1994 and references therein), and it can be taken to introduce children to the social and linguistic constraints that operate in the adult language. Note, however, that in this study the caregiver data cannot be taken to represent the exact input that children receive, since one cannot be sure that in the experimental situation caregivers, being aware of the social impact of the variants, would not modify their linguistic behavior in order to be "politically correct". In other words, one cannot be sure whether they use semantic agreement deliberately or

⁴Not all caregivers could participate in the experimentation. Nevertheless, the data sample collected from 21 adult speakers is rather large and allows me to compare children's and caregivers' agreement preferences.

spontaneously. Nevertheless, caregivers' production can be taken to represent the general pattern, since, as pointed out in Chapter 2, semantic agreement is taking over in Contemporary Russian with regard to hybrids referring to females and it is also common for female names in *-ok/-ik* on the verbs in past (cf. the discussion in Sections 2.2, 2.4.3, and 2.4.4).

There is another interesting issue that appears in Table 7.4. Caregivers also use the semantic rule more often with female names (97.8%) than with hybrids (78.5%). The same is not true of the children, who produced even fewer feminine agreement forms for female names (14.3%) than for hybrids (18.8%). However, as I am going to demonstrate below, the situation changes for older children.

Finally, syntactic masculine agreement is also attested in the caregivers' production, but the percentage is rather low: 2.2% for female names and 21.5% for hybrids. Interestingly, on the individual level, as shown in Tables 10.11 and 10.14 in Appendix II, only two adults use masculine agreement on a par with feminine for female names, while 16 (i.e. more than half) do so for hybrids referring to females. All other adult speakers, i.e. 19 in the case of female names and five in the case of hybrids, produce only feminine agreement forms with these noun types. These results suggest that semantic agreement is almost categorical in the case of female names, while there is a lot of variability with regard to hybrids. Note that children's production is also less variable for female names than for hybrids, but unlike the caregivers, children seem to be biased towards masculine.

In order to follow the development of the variable pattern during the preschool age, the data from twelve older children (age 5;1-6;5) are considered in Table 7.5. Individual results are shown in Tables 10.12 and 10.15 in Appendix II. Several important points arise from the data in Table 7.5. First, older children's production is variable, i.e. similarly to the younger children in the main study, they are able to use syntactic masculine agreement as well as semantic feminine. On the individual level, five children (i.e. approximately half) used variable agreement with female names and nine with hybrids. Second, the overall results suggest that older children use semantic agreement with female names rather frequently (61.9%), but the percentage of semantic agreement forms with hybrids is considerably lower (26.7%). In the case of hybrids they show a strong preference for syntactic masculine agreement (73.3%). Thus, not only do five- and six-year-olds appear to employ the semantic principle more often than two- and three-year-olds in the main study, they also start using it more consistently with female names similarly to the adult speakers.

Table 7.5: Children's overall agreement production with female names in *-ok/-ik* and hybrid nouns referring to females (12 children, age 5;1-6;5)

	M	F	Total
	N(%)	N(%)	(100%)
Female names in <i>-ok/-ik</i>	37 (38.1)	60 (61.9)	97
Hybrids referring to females	99 (73.3)	36 (26.7)	135

To summarize, both two- and three-year-olds as well as five- and six-year-olds demonstrate variable use of syntactic masculine and semantic feminine agreement forms with female names in *-ok/-ik* and hybrids referring to females. On the other hand, caregivers demonstrate variation only in the case of hybrids. Children and their caregivers also differ in their ability to use the semantic principle with these nouns. Two- and three-year-olds are very prone to construe agreement on the basis of morphology, while the caregivers do it on the basis of semantics. Looking at older children we see that the pattern is in the process of change, since the semantic principle has started being used at an intermediate level by five- and six-year-olds. Thus, older children comprise a mid-level group, which is different from younger children and presents a closer match to adults with regard to the acquisition of gender with both female names and hybrids referring to females. In what follows I account for these findings in light of the specific hypotheses and predictions related to the acquisition of constraints on variable input.

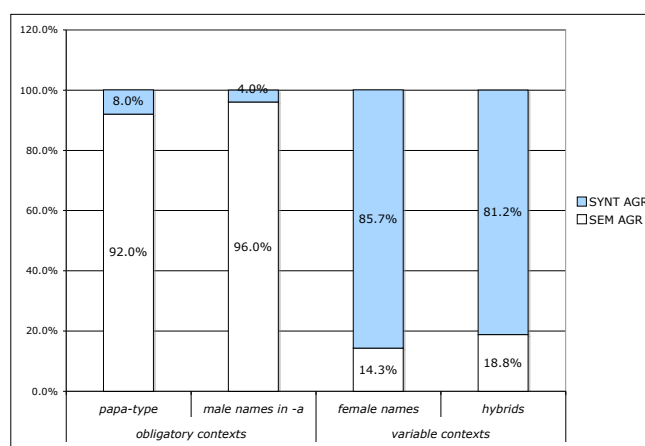
7.4 Acquisition of the semantic rule: Obligatory vs. variable contexts

The results from the child's data discussed in the previous section are consistent with Hypothesis I, which suggested that the acquisition of the semantic principle in variable contexts would be delayed. Both Prediction 1a and Prediction 1b are borne out. In this section I compare children's agreement production for hybrids referring to females and female names in *-ok/-ik* to their agreement production with *papa*-type nouns and male names in *-a* and show that with regard to the semantic rule, children demonstrate different agreement behavior for the two noun classes, so that it is used considerably less often for hybrids and female names than for *papa*-type nouns and male names in *-a*. Thus, Hypothesis II, which suggested that the semantic principle might be acquired simultaneously for all noun classes, is not supported.

Recall from the results presented in the previous section that I concluded

that the semantic criterion, i.e. the sex of the referent, is not determinant in eliciting gender agreement with female names and hybrids referring to females. On the contrary, children's agreement behavior seems to be dominated by the morphological principle, so that masculine agreement forms are used considerably more often than feminine: compare 85.7% vs. 14.3% for female names and 81.2% vs. 18.8% for hybrids. The reverse pattern was reported in Chapter 6 for nouns used in obligatory contexts (i.e. *papa*-type nouns and male names in *-a*), where on the overall level the proportions of semantic (masculine) and syntactic (feminine) agreement are 92.0% vs. 8.0% for *papa*-type nouns and 96.0% vs. 4.0% for male names in *-a* respectively. The results in Tables 6.1 (cf. Chapter 6, Section 6.3) and 7.3 are displayed in Figure 7.1, providing a comparison between children's agreement behavior in variable and obligatory contexts. It is clear in Figure 7.1 that the semantic procedure is very strong in obligatory contexts, since the percentage of masculine agreement is above 90% for both noun types; however, it is very weak in variable contexts where syntactic masculine agreement prevails over the semantic feminine.

Figure 7.1: The overall agreement distribution with nouns in obligatory and variable contexts (25 children, age 2;6-4;0)



Comparing the results of children's agreement production in obligatory vs. variable contexts reveals that the acquisition of the semantic principle is delayed in the case of hybrids referring to females and female names in *-ok/-ik* for whom both agreement forms are grammatical alternatives. I suggest

that the observed delay can be accounted for by the socio-cultural nature of the constraint on use of the semantic rule, which is found to be non-operative with these nouns in the speech of two- and three-year-olds. In other words, the fact that children do not consistently use the semantic principle in gender assignment with these nouns suggests that socially meaningful features have no linguistic significance to them at this age. Nevertheless, as shown in Chapter 6, they demonstrate full mastery of the grammatical constraint on use of the semantic rule with *papa*-type nouns and male names in *-a*. That is, children are aware of the male vs. female distinction expressed by the nouns, but they respect the dominance of the semantic principle only when the grammar requires it. Thus, on the one hand, children lack the ability to use the semantic rule to produce social meaning, but on the other hand, they are able to use it for grammatical purposes. These findings allow me to conclude that by age 4 children have gained grammatical competence but not socio-linguistic competence.

The observed time gap between the acquisition of the grammatical competence and socio-linguistic competence indicates children's early awareness of categorical vs. variable environments. It is also consistent with the claim that the categorical vs. extra-linguistic nature of the constraints that operate in language has an impact on the acquisition process, so that grammar-internal constraints are acquired before extra-linguistic ones. This is similar to previous findings on other morpho-syntactic variables, such as e.g. verbal *-s* in Buckie (Smith 2007), discussed in Chapter 3 Section 3.4.

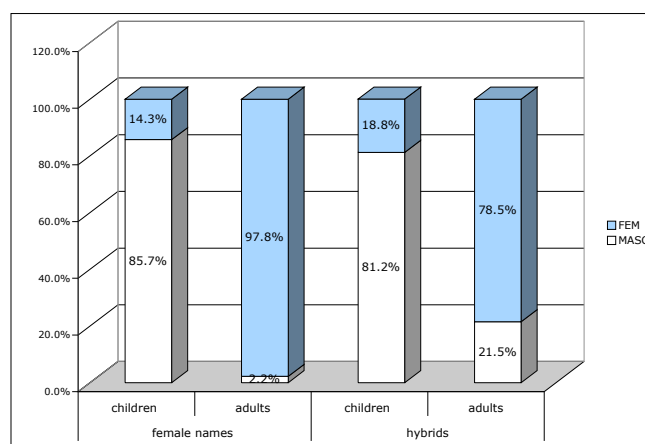
In the next section I take a closer look at the child and caregiver production which provides additional evidence for the claim that two- and three-year-olds lack socio-linguistic competence.

7.5 The distribution of agreement forms in child and caregiver data

So far I have discussed the evidence which confirms my first prediction that in variable contexts the semantic rule is used to a lesser degree than in obligatory contexts, since the constraint on its use is socially motivated. In addition, the results of child and adult production displayed in Tables 7.3 and 7.4 provide rich evidence for Hypothesis I, which also predicts that children use the semantic rule with these nouns less frequently than their caregivers. The overall results presented graphically in Figure 7.2 suggest that children aged 2;6-4;0 do not share the agreement pattern of their caregivers, who use semantic agreement predominantly with hybrids referring to females, i.e.

78.5% of the time, and with female names almost categorically, i.e. at a rate of 97.8%. In contrast to their caregivers, children use the semantic agreement seldom: at a rate of 14.3% for female names and 18.8% for hybrids.

Figure 7.2: Agreement between children and caregivers: female names in *-ok*, *-ik* and hybrids referring to females



It can be argued that this is a general pattern, since a similar contrast is documented at the individual level (see Tables 10.10 and 10.13 in Appendix II), where in the case of female names 17 children (out of 21 whose caregivers participated in the experimentation) used the morphological rule predominantly and in the case of hybrids referring to females 15 children did so.

Thus, it appears that socially motivated variability causes the children to initially rely on morphology rather than semantics. These findings suggest that for young children gender is a purely grammatical category and not a socio-linguistic phenomenon. The remaining question is: When and how do they come to match the norms of their parents?

Following Kerswill and Williams (2000:105), it can be predicted that this knowledge becomes available as a matter of socio-linguistic maturation “in a manner that involves a gradual increase in the number of styles that are perceived and treated in an adult way”. Recall from Section 3.4 that Smith (2007) has also suggested that caregivers can aid the children in this process by taking the role of “teachers of language”. This means that if two- and three-year-olds use syntactic masculine agreement predominantly, at a later

stage of development they should add another variant, i.e. semantic feminine agreement, for socio-cultural reasons. In the next section I show that this developmental prediction is borne out, since five- and six-year-olds demonstrate active improvement in their learning of socio-cultural constraints, which is especially clear with female names.

The beginning of this process is already noticeable between the ages of 2 and 3. In Figure 7.3 and Table 7.6 the comparison of the agreement rates between caregivers and children across individual hybrid nouns reveals that the highest rate of semantic (feminine) agreement in child production occurs for the noun *vrač* ‘physician’, i.e. 51.1%, which is also the closest match to the caregivers’ data, who used feminine agreement for this noun at a rate of 81.0%. This should not be surprising taking into account two facts. First, already very young children are familiar with physicians. In fact, people of this profession may be more familiar to them than librarians, policewomen, etc. Second, physician is typically a female occupation in Russia, which means that children may already have an association between a physician and a female person. Note that among adults this professional title has a very strong female stereotype (cf. Kapatsinski 2006). Given these facts, it is possible that children gain socio-cultural maturity with the noun *vrač* ‘physician’ earlier than with the other hybrids used in the test.⁵

⁵In Čukovskij’s (1965:439) notes from the diaries of some Russian children aged 2 to 5 I found anecdotal evidence suggesting that children can form associations between a professional title and a particular sex based on their real life experience. In the adult-child conversation below you can see that the child associates a physician with a female, and engineer with a male (the exact age of this child is unknown):

Adult: *Kem ty xočeš byt’, kogda vyrasteš?*

‘What do you want to be when you grow up?’

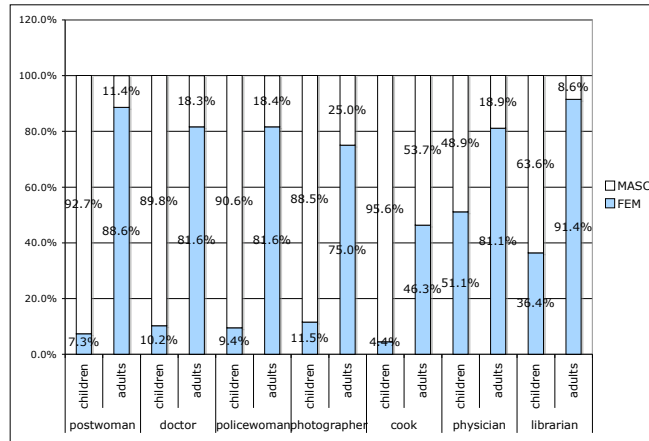
Child: *Esli ja vyrastu tetej, budu vračem. A vyrastu djadej - inženerom.*

‘If I grow up and become a woman, I will be a physician. And if I become a man, I will be an engineer.’

Table 7.6: Agreement distribution across individual hybrid nouns referring to females in child and caregiver data (25 children, age 2;6-4;0 and 21 adult speakers). Number of tokens

	<i>počtal'om</i> 'postwoman'	<i>doktor</i> 'doctor'	<i>milicioner</i> 'policewoman'	<i>fotograf</i> 'fotographer'	<i>povar</i> 'cook'	<i>vrac̃</i> 'physician'	<i>bibliotekar'</i> 'librarian'	Total
Children	38/3 M/F	44/5 M/F	48/5 M/F	23/3 M/F	43/2 M/F	22/23 M/F	28/16 M/F	246/57 M/F
Caregivers	4/29	8/36	7/38	6/16	20/17	5/28	3/29	53/193

Figure 7.3: Agreement between children and caregivers: Individual hybrid nouns referring to females

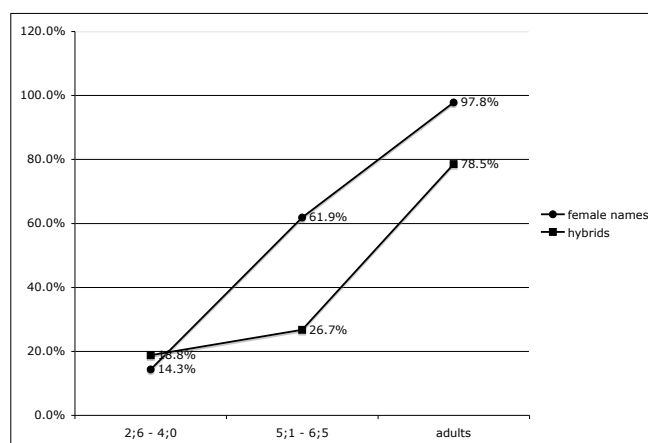


7.6 When does semantic agreement gain dominance in variable contexts?

In this section I compare the usage of semantic agreement with female names and hybrids in younger and older children. It has been predicted that as children grow older, they mature socially, hence they start using the semantic rule more frequently.

The predicted developmental trajectory stands out clearly in Figure 7.4 where the rates of semantic (feminine) agreement are compared for younger and older children for two noun groups (the numbers and percentages are given in Tables 7.3, 7.4, and 7.5).

Figure 7.4: Semantic agreement between younger children (25 children, age 2;6-4;0), older children (12 children, age 5;1-6;5), and adults (21 adult speakers): overall results for hybrids referring to females and female names in *-ok/-ik*



It is clear in Figure 7.4 that five- and six-year-olds use the semantic rule considerably more often with female names than two- and three-year-olds. The rates of semantic agreement increase for hybrids as well, although the difference between the two age groups is not big. Thus, younger and older children differ in their ability to use the semantic rule. These findings suggest that between the ages of 5 and 6 children's choice of the agreement forms with these nouns is more and more affected by the referents' biological sex. Children's hesitations and spontaneous self-corrections are particularly informative in this respect. In the speech of one child (Dusya 6;5) I found that she could spontaneously produce the masculine form with a female name and then immediately change the agreement for feminine:

- (6) *prosnulsja* +// *prosnulas'* (Dusya 6;5)
 wake_{PST.refl.M} up *wake_{PST.refl.F} up*
 ' (Valek) woke up.'

Thus, it can be argued that five- and six-year-olds are aware of the social weight of the semantic rule and more generally as children grow older they are able to integrate their socio-cultural knowledge into the linguistic system.

As discussed in Section 7.5, there are important differences between the

caregivers' use of semantic agreement with female names vs. hybrids. That is, caregivers employ semantic agreement predominantly with hybrids (at the rate of 78.5%) and near-categorically with female names (at the rate of 97.8%). Given this, it is important to know whether there are differences between the acquisition of the semantic principle with female names vs. hybrids in children's production. In Figure 7.5 I demonstrate the child pattern for female names, and in Figure 7.6 for hybrids. In both figures the data from younger children are divided into three age groups in order to make the acquisition process more explicit. Numbers and percentages for younger children across the three age groups are illustrated in Table 7.7. The data from older children in these figures are taken to represent the next stage of development (these data were presented in Table 7.5).

Table 7.7: Agreement production for hybrids referring to females and female names in across three age groups (25 children, age 2;6-4;0)

Age group	hybrids		female names	
	M	F	M	F
	N (%)	N (%)	N (%)	N (%)
2;6 - 3;0	46 (73.0)	17 (27.0)	45 (93.7)	3(6.3)
3;1 - 3;3	50 (73.5)	18 (26.5)	52 (85.2)	9 (14.8)
3;6 - 4;0	150 (87.2)	22 (12.8)	94 (82.5)	20(17.5)

With regard to the use of the semantic principle with female names vs. hybrids two entirely different pictures emerge: while children's use of the semantic rule with female names is in a steady climb across the four stages, it stays at approximately the same level for hybrids. Thus, on the one hand we see that social maturity seems to be greater for five- and six-year-olds than for younger children, as demonstrated by the developmental data for female names, and yet the situation is very unstable, as the pattern for hybrids indicates. Why should these two noun types develop differently? I suggest that there are two facts that may be responsible here. First, female names refer to females only, while hybrids can in general refer to a female, a male, or a human being irrespective of sex. Second, since the caregiver pattern for female names is found to be almost categorical, it can be hypothesized that the actual input for these nouns is more consistent than the input for hybrids, therefore children may improve with female names faster. This line of reasoning will be developed further in Chapter 9.

Figure 7.5: Children’s agreement preferences with female names in *-ok/-ik* expressed as a percentage for four age groups: 2;6-3;0, 3;1-3;3, 3;6-4;0, and 5;1-6;5

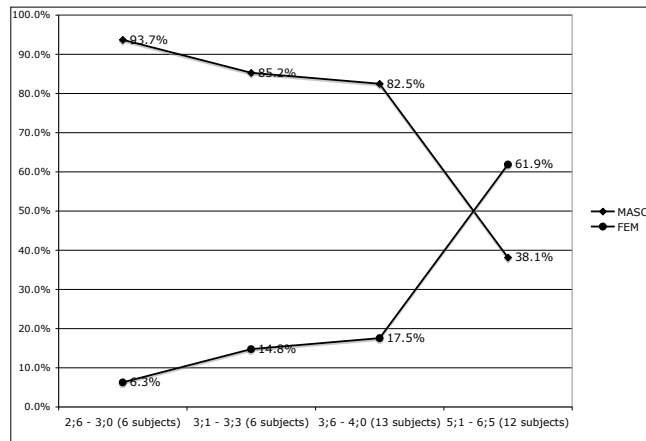
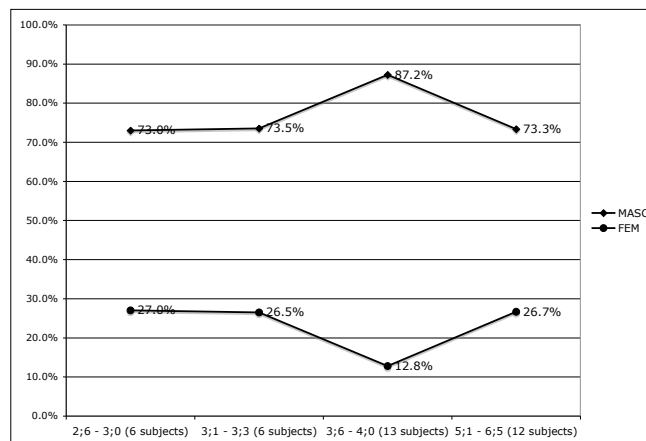


Figure 7.6: Children’s agreement preferences with hybrid nouns referring to females expressed as a percentage for four age groups: 2;6-3;0, 3;1-3;3, 3;6-4;0, and 5;1-6;5



With regard to Figure 7.6, it is rather unexpected the percentage of feminine forms with hybrids referring to females drops to 12.8% in the third age group, 3;6-4;0. Note that with female names it increases as compared to the second age groups from 14.8% to 17.5%. At the individual level it appears that 12 out of 13 children produced masculine forms with hybrids referring to females predominantly (see Table 10.13 Appendix II). In Table 7.7 above it is also clear that the sample produced by the children in the third age group is three times larger as compared to the samples of the younger children. There was just one child (Tolya 3;6) in this age group, who was equally disposed to use both masculine and feminine (4/4). Thus, the data appear appear to be very similar across the individual children in this age group. Hence their bias towards masculine with this subtype of nouns may be due to some other factors. One of them may be related to the non-categorical status of the semantic rule with this subtype, which is more variable than with female names that denote only females. Another possibility may be the exact input to these children, yet I have no relevant information available here. This has to be left open to further research.

Finally, the reader may recall that the differences in the acquisition of gender with proper names vs. common nouns have been also reported in Chapter 6. There I found that the semantic rule was acquired somewhat easier and earlier for male names in *-a* than for the low-frequency common nouns. This asymmetry was attributed to the differences in the semantic representation of proper names vs. common nouns that children seem to be sensitive to. More specifically, I proposed that the lack of additional semantic content in the representation of proper names may facilitate the acquisition of gender for this subtype of nouns. As pointed out in Chapter 6, this can be taken as an indication that proper names form a discrete class distinct from other nouns. The differences between female names in *-ok/-ik* and hybrids referring to females discussed above, being somewhat parallel to those between male names in *-a* and male kinship terms, still seem to bear on different issues. As mentioned above, these subtypes of nouns differ with regard to the qualitative and quantitative properties of the input, such as explicitness and consistency of the information necessary for the establishment of the semantic rule. I return to this issue in Chapter 9.

7.7 Summary and conclusion

Most importantly in this chapter we have seen that the course of acquisition of the semantic principle is different in variable vs. obligatory contexts. In variable contexts the observed course presents a much slower progression from

formal-to-semantic as compared to the obligatory contexts. The semantic principle appears to be a less favorable alternative even in the late pre-school years, as has been shown for hybrids referring to females. For female names its dominance may be established between the ages of 6 and 7. More evidence from older children is needed to make more specific conclusions about the exact time when the semantic rule becomes dominant for these nouns.

We have also seen that children's variation pattern is the opposite of that of their caregivers, who as expected use semantic agreement predominantly. Even though this is not the exact input the children in this study receive, this may be seen as an indication that children do not intend to match adults' behavior. More specifically, taking into account the shift mentioned above it can be concluded that in the course of gender acquisition children assume specific grammatical hypotheses, which are based on formal gender properties and which cause initial deviation from the adult grammar. Note that these hypotheses seem to be the same for the two subcategories of nouns I have looked at so far, i.e. masculine nouns in the second declension and hybrids and female names; yet the course of acquisition for them is different. The delay in the acquisition of the semantic rule for hybrids referring to females and female names in *-ok/-ik* has been taken as an indication that young children lack socio-linguistic competence. In other words, they are not aware of the social weight of the semantic rule; therefore they entertain the initial form-based hypothesis longer here. This finding also suggests that for young children gender is a grammatical category and not a socio-linguistic phenomenon.

The comparison of children's agreement production for the nouns in obligatory vs. variable contexts also supports the claim that the children in this study have gained grammatical but not socio-linguistic competence, as they use the semantic rule in the target-consistent way in variable contexts but are not prone to apply it in obligatory contexts. This also supports the idea that children are sensitive to categorical vs. variable environments and even further that they pay attention to the notion of class and generalize within it. On the other hand, the idea that children having acquired the semantic rule with one subcategory of nouns would immediately extend this knowledge to the other subcategory does not get support from the empirical data. In terms of input, it appears that ambiguous but consistent input in the case of *papa*-type nouns and male names in *-a* causes them to use the semantic rule predominantly; on the other hand, ambiguous but inconsistent input causes a delay for the establishment of the semantic principle in the case of hybrids referring to females and female names in *-ok/-ik*. Therefore, the role of input in the establishment of the gender category will receive special attention in Chapter 9 where I make an attempt to account for the course of acquisition

across all noun classes based on the qualitative/quantitative differences in the input.

In the next chapter I investigate the acquisition of gender with double gender nouns, a special class, which shares some properties of *papa*-type nouns and some of the hybrids. Its acquisition is thus expected to be problematic, yet children appear to have surprisingly little difficulty with them.

Chapter 8

Referential gender

8.1 Introduction

The last subcategory of nouns that I investigate in this dissertation is often referred to as nouns of common or double gender due to the fact that their gender value varies depending on the sex of the referent: it is masculine when the referent is a male (as in (1-a)) and feminine when it is a female (as in (1-b)).

- (1) a. Et-a plaksa revel-a vsju noč. (female referent)
this_F crybaby howl_{PST.F} all night
'This crybaby howled all night.'
- b. Etot plaksa revel vsju noč. (male referent)
this_M crybaby howl_{PST.M} all night

As outlined in Chapter 2, double gender nouns have no inherent gender. Their gender, often defined as referential gender (cf. Dahl 2000), must correspond to the sex of the individual they refer to. This means that their gender can only be determined in particular discourse situations. This chapter is thus concerned with children's ability to establish the gender of a noun via identification with a particular individual on each occasion when the noun is used. More specifically, the purpose of this chapter is to find out whether children attribute a decisive role to semantics in gender assignment with double gender nouns. Given that their form is typical of feminine nouns, special attention will be paid to the cases where the form of the noun does not correspond to the sex of the referent, i.e. when the noun refers to a male.

The experimental evidence presented in this chapter shows that children's choice of agreement with double gender nouns has referential justification already at an early stage, i.e. it corresponds to the sex of a referent: masculine

agreement is used productively and predominantly when the referent is a male and feminine when it is a female. It also shows that the same lexical item is attributed masculine gender when applied to a male and feminine when applied to a female.

The evidence presented in this chapter also provides support for the idea that children acquire the semantic principle for each type of noun individually, as their behavior for the nouns of double gender appears to be different from what has been observed for hybrids in the previous chapter. Moreover, children treat double gender nouns differently from hybrids when they are used non-generically and lack a sex specification. In particular, I will show that children use different agreement strategies for the two subcategories of nouns.

The chapter has the following organization. In Section 8.2 I formulate the predictions based on some evidence from previous acquisition research as well as certain theoretical considerations. The results of the experimentation are presented in Section 8.3. Section 8.3.1 presents the results of Experiment 8 on double gender nouns where the referents' biological sex was specified. Section 8.3.2 presents the results of Experiment 7 on double gender nouns and Experiment 5 on hybrids where the referents' biological sex was unspecified. In Section 8.4 the main findings are discussed with regard to the main hypotheses and predictions. Section 8.5 provides a summary of the chapter.

8.2 Hypotheses and predictions

In this section I discuss some predictions regarding children's ability to establish the correspondence between the sex of a referent and the gender of a double gender noun. In terms of gender assignment, the task of acquiring the referential gender seems to be rather complex: children have to realize that one lexical form, whose morphological properties suggest feminine, is in fact used to represent two gender values, feminine and masculine. Children have to make an assumption regarding which principle is determinant in gender assignment for these nouns: morphological or semantic. Given that the morphology of these nouns is clear and consistent, children, who are argued to be sensitive to morphological criteria early, can hypothesize that the gender of these nouns is derived by means of morphology, irrespective of the referent's sex. In this case they would erroneously assign feminine gender to nouns used in reference to males, since double gender nouns end in *-a* in nominative singular and belong to declension II, where the majority of feminine nouns are found in Russian. Only if children hypothesize that semantics/sex-based distinctions play a role will they be able to establish

correct gender representations. Clearly, in the case of a male referent where semantics and morphology are in conflict, the evidence that the semantic rule has been used will be most straightforward. If children know that the gender value of these nouns is referentially determined, they must be successful in establishing the correspondence between a male individual and the gender of a noun. They should have no difficulty doing so even though the form of the noun is misleading. In the case of a female referent the evidence that the semantic rule has been used cannot be conclusive, as semantics and morphology overlap. Yet, I suggest that if masculine and feminine are used consistently with an appropriate referent it can be an indication that the semantic rule is used in both cases, since after all, in order to be able to use feminine, the child has to make sure that the referent is not a male.

As pointed out in Chapter 3, very little is known about how children behave with regard to double gender nouns. From Gvozdev's (1961) and Popova's (1973) findings it appears that by approximately the age of 3, children have a clear idea how masculine and feminine nouns should look or sound, as they are able to assign masculine to nouns ending in a hard consonant and feminine to nouns ending in *-a*. Given children's reconstructions like *papa bjak* instead of *papa bjaka* 'daddy is a baddie' (cf. Čukovskij 1965:375), it is also clear that at some point of development children realize how a noun that refers to a male should look, hence they do not feel like using a 'feminine looking' noun in reference to a male. These reconstructions also indicate the strong impact that formal properties have on children's gender representations. However, previous research has not established whether children, despite the formal properties of double gender nouns, can make correct predictions of their gender in the case of a male. In other words, we do not know whether children can establish a correspondence between the sex of a specific individual and the noun, and whether they know that the same lexical item is assigned masculine gender when applied to a male and feminine when applied to a female. It is thus necessary to examine children's behavior with regard to the constructions presented in (1), where gender has referential justification.¹ For this purpose I designed Experiment 8 described in Chapter 5 Section 5.13. In order to be able to make a comparison between

¹In Chapter 2 it was mentioned that in copular constructions, where double gender nouns appear as predicates, they are not used referentially, e.g. *On izvestn-yj/-aja lakomka* 'He is a known_{M/F} gourmand'. There is thus no strict correspondence between the sex of the referent and the gender of a noun. Gender manifestations expressed by attributive adjectives can vary as you can see in the example above, where *lakomka* 'gourmand' is applied to a male. As I am interested in how children acquire referential gender, copular constructions are left aside in this study. They should be the topic of a separate investigation.

double gender nouns and hybrids I decided to focus on noun-verb (subject-predicate) agreement. The reader may recall from Chapter 7 that children were not willing to assign two genders to a hybrid noun: children's non-application of the semantic rule in reference to females was interpreted as a result of their not taking into account the semantic properties of the referent, instead giving preference to the morphological rule. It is thus especially interesting whether children can assign two genders equally well to the same lexical form. Therefore two of the nonce words used in the experimentation were tested twice: once they referred to a male and another time to a female in two different stories.

To summarize, I will be concerned with the following alternative hypotheses:

Hypothesis I. Children lack knowledge of referential gender.

Prediction 1. If children are not aware that the gender of a double gender noun must correspond to the sex of the referent in a particular situation, they should assign feminine gender to these nouns irrespective of the referent's sex. Consequently, erroneous feminine agreement will be prevalent for males.

Hypothesis II. Children have knowledge of referential gender.

Prediction 2. If children are aware that the gender of a double gender noun must correspond to the sex of the referent in a particular situation, then in the case of a male referent they should use masculine agreement predominantly. Moreover the accuracy rate should be equally high for males as for females.

Prediction 3. Since the situation with regard to the gender criteria is not equally balanced in the case of males vs. females we can expect a particular development. More specifically, since in the case of a male referent the morphological properties of a noun are misleading, younger children, i.e. two-year-olds, may perform worse than older children. In terms of agreement, erroneous feminine agreement can occur more often in the speech of younger children than in the speech of older children. At the same time, both younger and older children should perform equally well in the case of a female referent.

In addition the role of the referent's sex identity and the status of double gender nouns more generally has been tested in non-generic contexts where the sex of a referent was not specified. Experiment 7 (Section 5.12 Chapter 5) aimed to explore children agreement strategies with double gender nouns used in reference to specific individuals whose sex was inaccessible. This situation is rather complex, as sex is a governing factor for double gender nouns used in a particular discourse situation in reference to a specific individual. Given that children are equally disposed to choose either a male or a female they can assign gender arbitrarily, i.e. they can be expected to use masculine

and feminine randomly. Alternatively, they can be expected to prioritize morphology and produce only feminine forms. Still another possibility would be to use their own sex as the decisive factor in gender assignment.

Although Experiment 7 was not carried out with the children's caregivers, four adults (two females and two males) were tested after the experimentation was done. As discussed in Section 8.3.2 below, these adults used their own sex for gender assignment in this experimental situation, i.e. females gave 100% feminine responses and males gave 100% masculine responses. This observation will have interesting implications for the child data, yet due to the drawbacks in the experimental procedure I am not going to make strong claims here.

Children's agreement behavior with double gender nouns used non-generically and without sex specification will be further compared to their behavior with hybrids used in the identical situation. Hybrids used non-generically in reference to individuals without sex specification were tested in Experiment 5 (cf. Section 5.10 Chapter 5). In contrast to double gender nouns, biological sex is not a categorical factor for hybrids. Therefore, children can be expected to use syntactic (masculine) agreement predominantly here, as it is admissible for the referents of both sexes in the adult language (note that Experiment 5 was also tested with children's caregivers). The comparison of the results of Experiments 7 and 5 will reveal a number of important differences in children's behavior with double gender nouns vs. hybrids, which will be shown to support the idea that these two noun classes are treated differently by children.

8.3 Results

8.3.1 Double gender nouns - referent present

The results reported in this section were obtained in Experiment 8, which was designed to elicit noun-verb agreement with double gender nouns when the sex of their referents was accessible (see Chapter 5 Section 5.13). Individual results are presented in Tables 10.18 and 10.19 in Appendix II. When counting the data I excluded all unclear cases and counted every occurrence of a verb in past tense with or without a controller. Similarly to the other tests in my study, the majority of the structures produced by children in this test had covert subjects (i.e. 595 out of 603 totally). Only eight structures had overt subjects (five with a male referent and three with a female referent), all eight were target-consistent in terms of agreement which corresponded to the sex of the referent. As discussed in Chapter 7 Section 7.3, pervasive

subject omission is not surprising, given that it was pragmatically motivated by the experimental context where the test nouns and their referents were always in the center of discourse. Some examples of children's responses with covert and overt controllers are given in (2) for a male referent and in (3) for a female referent.

- (2) a. poedala leg spat' (Katya 2;11)
 heavy eater lie_{PST.M} sleep
 'The heavy eater went to sleep.'
- b. sjela (Katya 2;11)
 eat_{PST.F}
 '(The heavy eater) ate (the cake).'
- Target structure: sjel_{PST.M}
- (3) a. umnjaša dala im kostočku (Vera 3;9)
 smarty pants give_{PST.F} they bone
 'The smarty pants gave them a bone.'
- b. pomyl (Vera 3;9)
 wash_{PST.M}
 '(The smarty pants) washed (the dog).'
- Target structure: pomyla_{PST.F}
- c. vyterla (Vera 3;9)
 dry_{PST.F}
 '(The smarty pants) dried (the dog).'

With regard to the examples in (2) and (3) it should be noted that the same child could use correct semantic agreement, as in (2-a), (3-a) and (3-c), as well as erroneous syntactic agreement, as in (2-b) and (3-b), for the same referent, i.e. with regard to one picture in the story the agreement was correct but for another it could be wrong. It is most striking that children could err even when both criteria were pointing towards the same gender, as e.g. in (3-b). Agreement errors with double gender nouns used with reference to females (i.e. in the absence of a morpho-semantic mismatch) are rather unexpected and I return to them below.

Eighteen tokens containing pronoun-verb forms were excluded from the data since my main concern here was verbal agreement and its correspondence to the sex of the referent. Nevertheless, it should be noted that pronominal forms were used appropriately, i.e. *on* 'he' for males and *ona* 'she' for females, together with the matching verbal forms:

- (4) a. šljapu on nadel (Lyuba 3;3)
 hat he put_{pst.M} on
 ‘He put on a hat.’ (male)
- b. ona kisku požalela (Olya 2;7)
 she cat be_{pst.F} sorry
 ‘She was sorry for a cat.’ female

Erroneous use of the pronoun *ona* ‘she’ occurred three times in the speech of one child (Kolya 3;1). Nothing can be said about his knowledge of the correct form *on* ‘he’, as there are no relevant examples. With regard to verbal agreement, which I discuss below, this child was not target-like either. In fact he is one of those children who made errors not only when the noun applied to a male, but surprisingly also when it applied to a female.

The results of children’s verbal agreement production are presented in Table 8.1, which shows the overall distribution of agreement forms used with double gender nouns in reference to males and females. It is clear in Table 8.1 that children are very successful and nearly target-like in attributing masculine to double gender nouns when they are applied to males. The accuracy rate of 91.6% suggests that children do not have problems determining the gender of these nouns, even though it was predicted not to be easy, since the formal properties of these nouns are misleading and from the point of semantics, such as biological sex, the nouns themselves contain no clues. This result is thus very striking and seems to indicate that children can establish gender referentially via identification with a particular (male) individual, hence it seems to be consistent with Prediction 2.

Table 8.1: Overall distribution of agreement forms with double gender nouns used with reference to males and females (25 children, age 2;6-4;0)

Referent	M N (%)	F N (%)	Total N (100%)
	correct	erroneous	
Male	349 (91.6)	32 (8.4)	381
	erroneous	correct	
Female	28 (12.6)	194 (87.4)	222

With regard to Table 8.1 a problem arises when looking at children’s production for a female referent. In this case children are not expected to make any errors, since both the noun’s morphology and the referent’s sex suggest feminine. In the case of a female referent, the overall number of 28 errors result in an accuracy rate of 87.4%, which is slightly lower than in the

presumably more problematic case of a male referent where the accuracy rate amounts to 91.6% totally. Although the comparison of the error rates shows no significance between male and female referents ($p=0.5171$, $p \leq 0.1$) it is still surprising that children do make errors in the case of a female referent and even more than in the case of a male referent. This finding thus has to be investigated further.

Before I investigate this finding at the individual level, I would like to compare children's agreement production for double gender nouns referring to females in Table 8.1 to their production for hybrids referring to males previously presented in Table 7.3 Chapter 7. Although both cases are presumably unproblematic, since semantic and morphological features overlap, errors still occur. As demonstrated in Table 8.2, the number of errors are rather marginal in the case of hybrids (i.e. 3.1%), while it slightly exceeds the experimental error-margin of 10% in the case of double gender nouns, where it amounts to 12.6%. Thus, a unproblematic situation appears to be more problematic in the case of double gender nouns than in the case of hybrids.

Table 8.2: Overall distribution of agreement forms with double gender nouns used with reference to females and hybrids used with reference to males (25 children, age 2;6-4;0)

	M N (%)	F N (%)	Total N (100%)
	correct	erroneous	
Hybrids (male referent)	251 (96.9)	8 (3.1)	259
	erroneous	correct	
Double gender nouns (female referent)	28 (12.6)	194 (87.4)	222

The 28 errors reported in Tables 8.1 and 8.1 for double gender nouns referring to females did not occur in the speech of one child, although some children made more errors than others. Altogether ten children produced erroneous masculine agreement for double gender nouns applied to females. Three of them were girls, who made five errors. This means that the majority of errors (i.e. 23) were made by seven boys. Among them there was one child (Vasya 3;3) who showed a tendency to overuse masculine forms. For double gender nouns he produced only masculine agreement irrespective of the referent's sex. Thus he is responsible for almost half of the errors that the boys made, i.e. 10 out of 23.² The other six boys made one or two

²Despite the overuse of masculine by this child, feminine forms are also attested in his production. Specifically, he could use feminine erroneously with *papa*-type nouns and male names in *-a* and optionally with female names in *-ok*, *-ik* (cf. Tables 10.3, 10.4, 10.5,

errors each. Given these results, I further distributed children's production between boys and girls to investigate whether other factors, e.g. the child's own biological sex, can play a role here.

As shown in Table 8.3, the proportion of correct responses for male and female referents is rather balanced and equally good in the girls' production, while it is considerably lower for females than for males in the boys' speech. Given this, one might suggest, that boys' agreement production is conditioned by their own sex, rather than by the sex of the referent presented in the picture. It is reasonable to ask whether the speaker's sex also affects children's, and in particular boys', agreement production in other tests. As I show below, the answer to this question is rather 'no'. Together with some other considerations, which I consider below, it seems rather unlikely that the sex factor affects the agreement production by boys.

Table 8.3: Agreement production for boys and girls with double gender nouns used in reference to males and females (25 children, age 2;6-4;0)

	Male referent		Female referent	
	M	F	M	F
	N (%)	N (%)	N (%)	N (%)
Boys (11 children)	142 (93.4)	10 (6.6)	23 (27.1)	62 (72.9)
Girls (14 children)	207(90.4)	22 (9.6)	5 (3.6)	132 (96.4)

If the sex factor plays a role, then it can also be expected that the percentage of correct responses given by boys is higher for masculine nouns in the second declension than the percentage of correct responses given by girls. Moreover, with regard to hybrids referring to females and female names in *-ok*, *-ik*, boys should use masculine agreement more often than girls. However, evidence presented in Table 8.4 suggests the opposite. Boys make considerably more errors than girls with masculine nouns in the second declension (i.e. 13.7% vs. 4.8% respectively) and use masculine forms less than girls for hybrids referring to females and female names in *-ok*, *-ik* (i.e. 28.3% and 20.2% vs. 12.6% and 10.8%). Thus, there is no indication that the boys' production could be conditioned by their own sex. Moreover, if sex were the reason, then why should it affect only boys in these cases? In Table 8.3, it is also puzzling why girls are nearly as good as boys in attributing masculine gender to these nouns and why they also make errors when the referent applies to a female. Given this, it is possible that the differences between boys and girls illustrated in Table 8.3 are motivated by factors other than the children's own sex. It may be, for example, that boys simply paid less

and 10.10 in Appendix II).

attention to the sex of the individual in the picture when they were being tested than the girls, who sometimes were also inattentive.

Interestingly, in his famous investigation of proper names vs. common nouns reviewed in Section 6.4.3, Macnamara (1982), being puzzled by the boys' poor responses relative to the girls', explored a number of possibilities, such as sex, age/level of development, and method. However, his further investigation revealed that none of these factors were important. He finally came to the conclusion that the difference between boys and girls was due to the fact that boys paid less attention than girls to the test.

Table 8.4: Agreement production for boys and girls across three noun classes (25 children, age 2;6-4;0)

	Girls		Boys	
	M	F	M	F
	N (%)	N (%)	N (%)	N (%)
Masc. nouns in Decl. II	758 (95.2)	38 (4.8)	499 (86.3)	79 (13.7)
Hybrids ref. to females	160 (87.4)	23 (12.6)	86 (71.7)	34(28.3)
Female names in <i>-ok/-ik</i>	124 (89.2)	15 (10.8)	67 (79.8)	17 (20.2)

In interpreting the data further I divided the overall results in Table 8.1 by individual nouns, as shown in Table 8.5. Recall that except the existing but presumably unfamiliar nouns *bedolaga* 'poor wretch' and *stiljaga* 'mod', the test items were derived by the author from the existing stems, so that they were also unfamiliar. Three important points emerge once the data are divided in this way. First, the proportions of correct responses in the column for males, which range between 85% and 100% for individual nouns, suggest that masculine gender of these nouns was derived from the sex of their referents. Second, the gender of these double gender nouns is contextually-determined. Masculine and feminine responses are found to be in complementary distribution: masculine agreement was used predominantly when the noun applied to a male and feminine agreement was used predominantly when the noun applied to a female. The accuracy rates range between 85% and 100% for a male referent and between 70% and 96% for a female. Third, children seem to be able to assign two genders to the same noun, as suggested by the accuracy rates of 92.3% vs. 70.0% for *umnjaša* 'smarty pants' and 88.9% vs. 85.2% for *pačkuxa* 'sloven'.

Table 8.5: Overall distribution of agreement forms with double gender nouns used with reference to males or females (25 children, age 2;6-4;0)

	Male referent			Female referent	
	M	F		M	F
	N (%)	N (%)		N (%)	N (%)
umnjaša	60 (92.3)	5 (7.7)	umnjaša	15 (30.0)	50 (70.0)
pačkuxa	48 (88.9)	6 (11.1)	pačkuxa	8 (14.8)	46 (85.2)
obižala	41 (100)	0 (0)	terjaxa	3 (5.7)	50 (94.3)
poedala	50 (84.7)	9 (15.3)	žalelka	2 (4.0)	48 (96.0)
<i>bedolaga</i>	71 (93.4)	5 (6.6)			
<i>stiljaga</i>	79 (91.9)	7 (8.1)			

The results presented in this section may not be totally conclusive, as it is not clear why boys were worse than the girls in their agreement production for a female referent. Yet, there is an indication that children are able to attribute gender to these nouns based on the sex of a specific individual. In the next section I discuss this finding in the context of the results presented in the previous chapters and show that children's ability to use the semantic rules differs depending on the noun class.

8.3.2 Double gender nouns vs. hybrids - referent absent

In this section I present the results of experiments 5 and 7 where the test nouns were used with reference to particular individuals, i.e. non-generically, but the sex of their referents was not provided. Individual results are illustrated in Tables 10.16 and 10.17 in Appendix II. Two points emerge in Table 8.6, which displays the overall agreement production for hybrids and double gender nouns. First, children employ different number of agreement possibilities for the two noun types: two (masculine and feminine) for hybrids but three (masculine, feminine and plural) for double gender nouns. Plural agreement was found in the speech of 11 children. As pointed out in Section 5.7, ungrammatical plural verb forms were sometimes used by the experimenter in the lead-in question in several tasks, e.g. *Čto *sdelali_{PL}* (target: *sdelal_{M,SG}*) *Trenya?* [What did *Trenya* do?]. However, only in Experiment 7 discussed here were children affected by plural agreement in the elicitation question and produced plural in response. These plural responses cannot be interpreted as imitation, since children who heard the verb *sdelali* 'made' in the lead-in question, used different stems in their answers, e.g. *sprjatali*

‘hid_{PL}’. Thus plural forms might be due to the unnaturalness of the experimental situation where the child had to refer to a specific individual whose sex was unknown. Second, there are robust differences in the proportion of masculine and feminine agreement. In the case of hybrids, children exhibit very high rates of masculine agreement, i.e. 95.8%, while feminine forms constitute 4.2% only and occur in the speech of only three children. On the other hand, in the case of double gender nouns, the distribution of masculine and feminine forms is rather balanced: 53.2% and 36.9% respectively. The rest are the structures containing plural agreement, i.e. 9.9%, which, as I said above, are found in the speech of 11 children.

Table 8.6: Overall agreement production with hybrid and double gender nouns whose referents are unspecified for sex (25 children, age 2;6-4;0)

	M	F	PL	Total
	N (%)	N (%)	N (%)	(100%)
double gender nouns	150 (53.2)	104 (36.9)	28 (9.9)	282
hybrids	366 (95.8)	16 (4.2)	0 (0)	382

In (5) and (6) I provide a few examples of the relevant agreement forms in Table 8.6. The structures in (5) present three agreement possibilities with different double gender nouns, while (6-a) and (6-b) are examples of masculine and feminine forms attested with hybrids.

- (5) a. a potom umnjaša glaza postroil nosik i rotik (Vera
and then smarty pants eye build_{PST.M} nose and mouth
3;9)
‘And then the smarty pants made a nose and a mouth.’
- b. sprjatali (Vera 3;9)
hide_{PST.PL}
‘(The decorator) hid (the presents).’
- c. postavila pod elku (Liza 3;9)
put_{PST.F} under fir tree
‘(The decorator) put (the presents) under the Christmas tree.’
- (6) a. a potom dvornik listočki tuda i povez (Roza
and then yard-keeper leaves there and carry_{PST.M}
3;10)
‘And the the yard-keeper (put) the leaves there and carried them.’
- b. sup grela (Dima 2;6)
soup warm up_{PST.F}
‘(The cook) warmed up the soup.’

Unfortunately, I only tested caregiver's agreement preferences with hybrids. With regard to double gender nouns the adult data comes from four native speakers, two females and two males, who were tested in Tromsø after the experimentation was completed. For hybrids, when their referent's sex was unknown, caregivers produced masculine agreement 100% of the time. Thus, children's production with hybrids, which amounts to 95.8% as shown in Table 8.6, is a very close match to their caregivers. However, children's behavior with regard to double gender nouns appears to be different from that of the adults. While children produced both masculine and feminine, adults used only one agreement type exclusively. Females always used feminine and males always masculine. It thus appears that the adults constructed agreement in accordance with their own sex. Given these results, I subdivided the data from children for boys and girls, as shown in Table 8.7.

Table 8.7: Agreement production for boys and girls with double gender nouns whose referents are unspecified for sex (25 children, age 2;6-4;0)

	M	F	Total
	N (%)	N (%)	(100%)
Boys (11 children)	72 (75.0)	24 (25.0)	96
Girls (14 children)	78 (49.4)	80 (50.6)	158

It is clear in Table 8.7 that girls' responses were random; they were equally disposed to use either masculine or feminine. Thus, their agreement behavior does not match the behavior of the adult females, who used feminine 100% of the time. Boys, on the other hand, show a tendency to overuse masculine, which is thus similar to the behavior of the adult males, whose responses were 100% masculine. To conclude, boys seem to be biased towards masculine when the sex of the referent is known and it is a female as well as when the sex of the referent is unknown. Girls, on the other hand, behave differently from boys as well as from the adults.

8.3.3 Summary of the results

With regard to gender assignment with double gender nouns whose referent is known, I find the following:

1. The overall results suggest that the usage of masculine forms for a male referent is highly (roughly 90%) adult-like.
2. The gender of double gender nouns corresponds to the sex of their referents approximately 90% of the time for males as well as for females.

3. Children can assign two genders to the same lexical item: masculine in the context of a male and feminine in the context of a female.
4. Children's production seems to be slightly worse in the case of female referents due to the boy's tendency to overuse masculine. The detailed investigation of the data shows that this tendency is only evident for double gender nouns whose referent is either known or unknown. It seems that the sex factor, i.e. that boys apply their own sex rather than the sex of the referent, cannot be totally excluded, since the adults use this strategy in the case when the referent is unknown. Note, however, that this strategy is not appropriate in the adult language when the referent is known.

With regard to gender assignment with double gender nouns and hybrids whose referent is unknown, I find the following:

1. Children behave differently with double gender nouns as compared to hybrids. Not surprisingly, there seems to be some confusion with regard to what agreement form should be used when the referent of a double gender nouns is unknown. Children produce masculine, feminine and ungrammatical plural forms. With regard to hybrids there is no hesitation or confusion. Masculine forms are used predominantly.
2. In the case of double gender nouns, boys appear to be a closer match to the adults, as they seem to use their own sex in gender assignment here. Girls, on the other hand, do not match the adults. They are equally disposed to use either masculine or feminine.

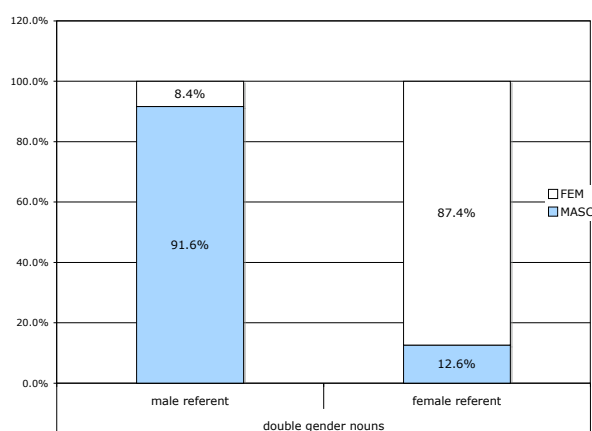
I now turn to the discussion of these results in the light of the specific hypotheses and predictions formulated in Section 8.2.

8.4 Knowledge of referential gender

With regard to the main question addressed in this chapter the results of the experiments show that children aged 2;6-4;0 are able to establish the correspondence between the sex of a particular individual and the gender of a double gender noun on each occasion of its use. This is consistent with Hypothesis II, which predicted that children would have knowledge of referential gender. Thus, Hypothesis I and Prediction 1 are not supported. On the other hand, Prediction 2 seems to be borne out, as children can be seen as largely using a semantic male-female distinction a guideline in the choice of agreement forms, as shown in Figure 8.1 (cf. the data in Table 8.1). The

graph representing children's agreement production for males shows that the referent's sex has been taken into consideration to a great extent, at a rate of 91.6%. This means that when deciding on the gender of a double gender noun, which does not have a sex distinction as part of its lexical representation, children pay attention to the sex of the referent, i.e. whether it is a male or not. Clearly, masculine and feminine forms are in complementary distribution for double gender nouns: 91.6% vs. 87.4% respectively. Still one cannot be certain that sex was the determining criterion in the case of females. Children could as well hypothesize that sex is crucial when the noun applies to a male but that morphology can be used otherwise. I suggest that more reliable evidence in this respect can be obtained in a comprehension task, where children may be asked about the sex of the individual based on the agreement forms they were introduced to. This is a topic of interest for future research.

Figure 8.1: The overall agreement distribution with double gender nouns used with reference to males and females (25 children, age 2;6-4;0)



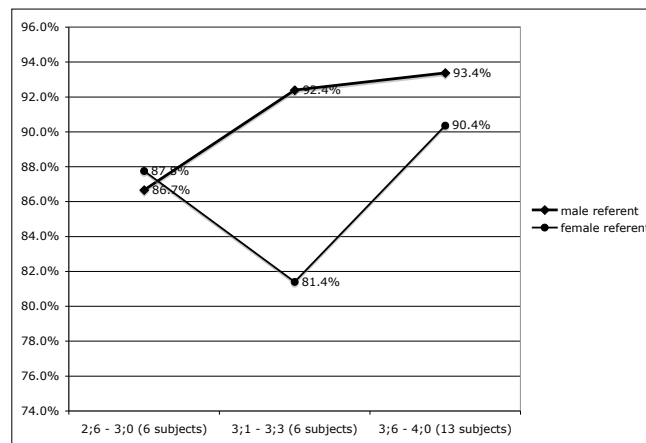
With regard to Prediction 3 we would expect younger children to experience more problems with double gender nouns than older children especially in the case of a male referent, where morphology is misleading. Table 8.8 and Figure 8.2 show that Prediction 3 is borne out. As predicted, two-year-olds make more errors than three-year-olds, and especially the children aged 3;6-4;0. The line graph in Figure 8.2 shows a clear increase of the accuracy rates for a male referent. The absence of a big jump towards the adult-like use

between the youngest and the oldest children suggests that the knowledge of referential gender may already be present in two-year-olds. High accuracy rates of 86.7% for a male referent and 87.8% for a female referent observed between the ages of 2;6 and 3;0 suggest that children's choice of agreement with double gender nouns has referential justification already at an early stage.

Table 8.8: Agreement distribution with double gender nouns used with reference to males and females across three age groups (25 children, age 2;6-4;0)

Age group	Male referent		Female referent	
	M	F	F	M
	N (%)	N (%)	N (%)	N (%)
2;6 - 3;0	78 (86.7)	12 (13.3)	43 (87.8)	6 (12.2)
3;1 - 3;3	85 (92.4)	7 (7.6)	48 (81.4)	11 (18.6)
3;6 - 3;10	169 (93.4)	12 (6.6)	103 (90.4)	11 (9.6)

Figure 8.2: Developmental curves of target-consistent agreement with double gender nouns across the three age groups (25 children, age 2;6-4;0)

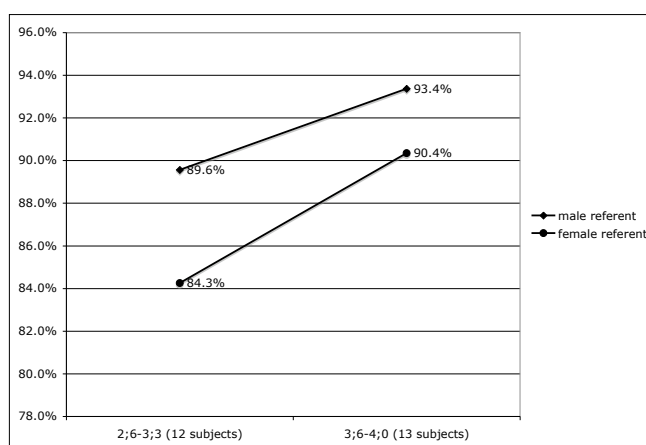


According to Figure 8.2, in the case of a female referent, there is a sudden drop of accuracy rates from 87.7% to 81.4% for the middle age group. Only two children (both boys) made errors in the middle age groups. Ten out of 11 errors totally occurred in the speech of one child (Vasya 3;3), who was previously mentioned in Section 8.3.1. Presumably, if the data of this child

are eliminated the drop will disappear. However, it seems more appropriate to collapse the data from the two younger groups into one. Recall that the same was proposed in Chapter 6 for the analysis of gender assignment with *papa*-type nouns and male names in *-a*.³

Once the data from the two younger groups are collapsed, as in Figure 8.3, we see a gradual and continuous developmental progress across the two groups. However, children's production with a female referent is still somewhat worse than with a male. I have as yet no clear explanation for this.⁴

Figure 8.3: The developmental curves of target-consistent agreement with double gender nouns across two age groups (25 children, age 2;6-4;0)



Thus, despite the complexity of the gender assignment task with double gender nouns, it can be argued that gender assignment is guided by semantic rules. When children's agreement behavior for double gender nouns is compared to their behavior for *papa*-type nouns and male names in *-a*, where

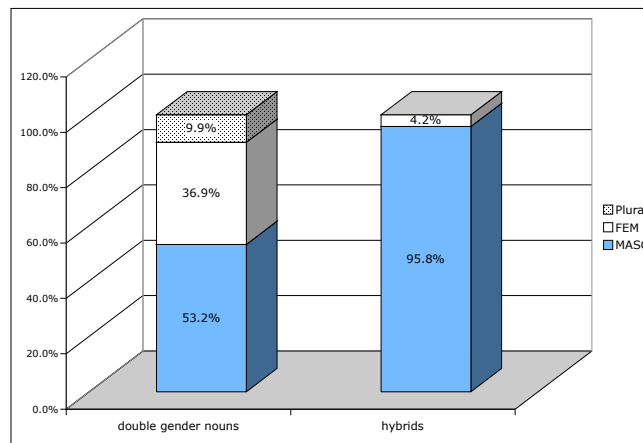
³Note that the children between the ages of 2;6 and 3;3 attended the same class in the day-care center and may thus have similar levels of linguistic development.

⁴Figures 10.1 and 10.2 in Appendix III show what the developmental pattern would look like if data of this one child are eliminated. Specifically, when the data are divided in three age groups the dip observed in Figure 8.2 for a female referent disappears and the line shows a sudden increase for the middle age group where the accuracy rate appears to be even higher than in the last age group: 98.0% vs. 90.4% respectively. With regard to a male referent, there is a gradual increase across the three age groups from 86.7% for the youngest children to 93.5% for the oldest. If the data are divided in two age groups (2;6-3;3 and 3;6-4;0), the accuracy rates raise from 88.9% to 93.5% for a male referent but they decline from 92.3% to 90.4% for a female referent.

the semantic rule was shown to be a dominant factor already at the age of 2;6, it is rather surprising that the phenomenon of double gender is almost as easy to deal with as lexical (inherent) gender. Why should it be so? An answer to this question will be proposed in Chapter 9, where I consider the role of input in discovering gender cues, and more specifically the consistency of representation and the obligatoriness of the semantic rule.

Furthermore, despite the morpho-semantic conflict, children are able to establish the correspondence between the sex of a referent and the gender of a noun and to assign two genders to the same lexical item: masculine when the noun is applied to a male and feminine when it is applied to a female. Thus, in the case of double gender nouns, children's behavior is absolutely different from their behavior for hybrids, where the referent's sex was disregarded in favor of the morphological criterion. That children's gender representations are different for double gender nouns as compared to hybrids is also clear from their agreement behavior when the sex of the referent is unknown. The differences in agreement patterns between hybrids and double gender nouns were illustrated in Table 8.6 Section 8.3.2. In this section they are expressed graphically in Figure 8.4. These differences may be an indication that hybrids and double gender nouns are treated differently by children.

Figure 8.4: The distribution of agreement forms with hybrid and double gender nouns with no sex specification (25 children, age 2;6-4;0)



The most striking observation with regard to Figure 8.4 is the following. In the case of hybrids the children are prone to construe agreement on the

basis of the noun's morphology, as 95.8% of their responses are masculine. On the other hand, in the case of double gender nouns only 36.9% of their responses are feminine, which suggests that children are not prone to apply the morphological rule here. In Table 8.7 we have also seen that in the girls' production the percentage of masculine responses was equal to the percentage of feminine responses: 49.4% vs. 50.6% respectively. The reason for this may be that girls were trying to guess the referent's sex being aware that in a particular discourse situation with reference to a specific individual the gender of a double gender noun must correspond to the sex of the referent. Recall that unlike girls, boys were biased towards masculine; 75.0% of their responses were masculine and 25.0% were feminine. The overuse of masculine documented in the speech of boys as well as male adults has been interpreted as a strategy to apply the speaker's own sex. The important observation is yet that the strong impact of formal properties is evident for hybrids but not for double gender nouns.

The overall picture in Figure 8.4 thus indicates that children treat hybrids and double gender nouns in different ways. It seems easy for them to take a decision in favor of syntactic masculine in the case of hybrids which corresponds to their form. However, they do not resort to the morphological strategy with double gender nouns, but try a number of possibilities including ungrammatical plural. This uncertainty I suggest may be taken as an indication of children's knowledge that the gender of these nouns should be contextually determined.

8.5 Summary of the chapter

In this chapter I examined children's knowledge of referential gender, i.e. their ability to establish the gender of a double gender noun via identification with a particular individual on each occasion when the noun is used. I concluded that two- and three-year-olds have knowledge of referential gender based on the complementary relationship of masculine and feminine agreement forms used with reference to males and females respectively and consistently by children. It is also important that the semantic rule was used productively and predominantly for nouns that applied to males and that the same lexical item was assigned two different genders depending on the referent's sex.

The results reported in this chapter show that the two subcategories of nouns that do not have inherent gender are treated differently by children. Sex of the referent is a dominant factor in the case of double gender nouns, while morphology tends to override semantics in the case of hybrids.

I have also shown that morphology does not have an impact on children's agreement preferences for double gender nouns used non-generically with no sex specification as it does with hybrids. It is thus possible that the sex of the referent may have some impact even when it is inaccessible as well as the speaker's own sex.

The next chapter summarizes and reconsidered the results of this experimental study in order to provide a more general account of the acquisition process in relation to gender. It also focuses on the implications of the results for the gender assignment theory.

Chapter 9

Wrapping things up

9.1 Introduction

I started my investigation by presenting gender as a grammatical category which poses a challenge for a first-language learner. First, this category may involve elaborate language-specific inflectional morphology, which represents grammar-internal linguistic knowledge. Second, it can require the understanding of semantic (cognitive) concepts related to meaning and reference, which is on the opposite extra-grammatical; it comes from the knowledge of the real world and must be integrated into the linguistic system. Russian is one of the many gender systems where the two factors are closely interrelated and where the child in order to attribute gender to a noun has to consider both of these factors and establish correct relationships between them. Thus, acquiring a gender system like that of Russian means acquiring two types of knowledge: grammatical and conceptual, and relating them to each other. Therefore, it seemed important to investigate how children organize nouns into genders in the course of acquisition.

In the previous chapters I have explored how children acquire gender of the exceptional nouns whose formal properties do not allow a correct prediction of their gender. My main concern has been children's ability to make use of the semantic concept of natural gender, i.e. to realize its grammatical function and integrate it into the language system. This process has been viewed as problematic as two of the subcategories of nouns under discussion are not sex-differentiable, namely hybrids and double gender nouns. Therefore it was important to examine whether children can establish a noun's gender via identification with an individual in a concrete discourse situation as well as their awareness of the fact that some nouns have one grammatical gender, while others can have two.

The experimental work has been done to show how and when the semantic concept is integrated into the gender system. The empirical facts about the acquisition of gender that I received through experimentation are summarized and reconsidered in this chapter in order to provide a more complete account of the gender acquisition process. The main finding, which shows that children acquire gender of each subcategory of noun individually, will be interpreted here as a result of children's being sensitive to specific input cues. The course of acquisition will thus be presented as a cue-based process where input is attributed an important role in discovering those cues.

In Section 9.2 I summarize the relevant for discussion results of the experiments which were presented and analyzed in Chapters 6, 7, and 8. In the following sections I provide a general account for the course of gender acquisition across various subcategories of nouns. In Section 9.3.1 I present the course of acquisition as a gradual change from formal to semantic. In Section 9.3.2 I present gender acquisition as a cue-based process. In Section 9.3.3 I attempt to account for the differences in the acquisition of the semantic rule across the noun classes. In Section 9.4 I reconsider Corbett's idea of a semantic core in light of the novel acquisition data. More specifically, I question the core function of the semantic criterion. Finally, in Section 9.5 based on the novel experimental findings and previous cross-linguistic observations I propose that gender acquisition may represent innately guided learning.

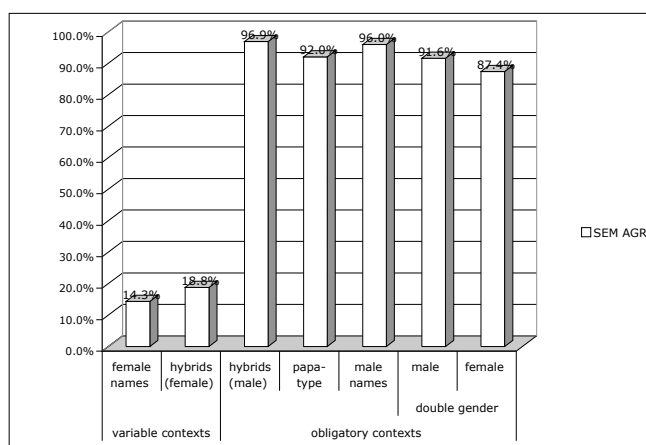
9.2 Overview

The goal of my experimental study was to determine the acquisition pattern of the semantic criterion for specific classes of Russian nouns. By looking at children's agreement behavior with each class in particular I discovered that the proportions of semantic agreement are highly differentiated across the noun classes so that children show much higher rates of semantic agreement for *papa*-type nouns, male names in *-a*, and double gender nouns than for female names in *-ok/-ik* and hybrids used in reference to females. Thus, the prediction formulated in Section 4.5, which suggested that as soon as the dominance of the semantic criterion is established with one subcategory of nouns this knowledge will be immediately extended to the others, is not supported. On the other hand, the differentiated use of the semantic rule is consistent with the idea that gender is acquired for each subcategory of nouns individually. Recall that according to the second part of this prediction, it is to be expected that the acquisition of three subtypes may be delayed, i.e. hybrids, female names and double gender nouns. Yet, the prediction is borne out for two of them, i.e. hybrids and female names, but not for double gender

nouns. I return to this result in Section 9.3.3.

The pattern that should be accounted for in a comprehensive analysis of the course of gender acquisition is summarized in Figure 9.1.

Figure 9.1: Semantic agreement across the noun classes. Overall results (25 children, age 2;6-4;0)



The following main points present particular interest for the discussion:

- The acquisition of the semantic rule is not simultaneous for all noun classes. Between the ages of 2;6 and 4;0 it appears to be operative (and dominant) for the nouns in obligatory contexts, while its acquisition is delayed for nouns in variable contexts.
- For *papa*-type nouns and male names in *-a* the semantic rule is operative and dominant already at the age of 2;6. According to the evidence in Chapter 6, it gains dominance gradually and most slowly for the low-frequency common nouns. For these nouns the mastery of the semantic rule is evident between the ages of 3;6 and 4;0, i.e. for the oldest children in the study.
- With regard to double gender nouns, where gender is contextually-determined, it corresponds to the referent's sex around 90% of the time in the child data for both male and female referents.
- Throughout the considered developmental period the semantic rule is used optionally and unwillingly for hybrids referring to females and fe-

male names in *-ok/-ik*. Thus in contrast to their caregivers, the children disregard the sex of a referent in gender assignment for these nouns. According to the additional evidence in Chapter 7, the integration of the semantic criterion proceeds faster for female names than for hybrids, as becomes evident in older children (age 5;1-6;5).

In the next section I attempt to provide a general account of the course of gender acquisition based on the results reviewed here.

9.3 The course of acquisition

The differences in the acquisition patterns presented in Figure 9.1 form the basis for two important claims. First, I argue that the acquisition of the semantic rule proceeds in the same way for different types of nouns. In other words, despite the differences there is a single underlying process. Second, I propose a cue-based account of the acquisition process and furthermore suggest that qualitative and quantitative differences in the input may have an impact on when the semantic rule is actually acquired.

9.3.1 The general developmental pattern

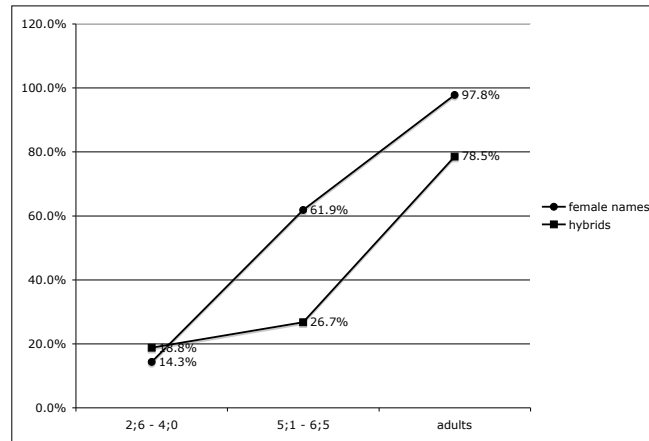
The general pattern that I argue for in this section is consistent with the previously made observation that children initially choose a formal basis for language development and that their language organization changes from formal to non-formal. I suggest that the differences in the acquisition patterns illustrated in Figure 9.1 can be taken as an indication that there is a single underlying process for all subcategories of nouns. Specifically, what we see is that a formal grammatical analysis based on the morphological properties of nouns is replaced by a non-formal analysis based of the semantic/sex-based properties of nouns. This change happens at different times for different nouns. Thus it is less apparent for nouns in obligatory contexts than for nouns in variable contexts.

Even though between the ages of 2;6 and 4;0 the change from formal to semantic is not very explicit for nouns in obligatory contexts, there are a number of indications that support this idea. Similar to other investigations of early child language this study shows that two- and three-year-olds are sensitive to the formal properties of nouns in gender assignment. Over-regularizations revealing this sensitivity were errors of attributing feminine gender to masculine nouns in Declension II or, as I often referred to them in this thesis, nouns ending in *-a*, i.e. *papa*-type nouns, male names in *-a* and double gender nouns. The impact of formal properties for these nouns

was shown to be stronger for younger children (between the ages of 2;6 and 3;3) than for older children (between the ages of 3;6 and 4;0). Moreover, the detailed analysis of the overregularization pattern in Chapter 6 revealed that it was also stronger for individual rare kinship terms such as *mužčina* ‘uncle/man’ and *junoša* ‘youth’ than for familiar words like e.g. *papa* ‘daddy’. Furthermore, given the asymmetry in children’s production with male kinship terms analyzed in Chapter 6, it is possible that at an earlier stage the amount of overgeneralizations would be larger, especially for low-frequency nouns like *mužčina* ‘uncle/man’ and *junoša* ‘youth’. In sum, although for nouns in obligatory contexts the change was visible already at the age of 2;6, we could still see the gradual transition towards the target-like use for the low-frequency nouns, in which case the complete mastery of the semantic rule (at a 90% level) was only at around age 3;6.

The transition from formal to non-formal was shown more explicitly for the nouns in variable contexts, especially for female names in *-ok/-ik*. In Chapter 7 I showed that children start out by using the morphological rule predominantly for hybrids referring to females and female names in *-ok/-ik* and only later in development (between the ages of 5 and 6 for female names and for hybrids presumably even later) they change their strategy towards the adult-like use where the semantic criterion is dominant. As illustrated in Figure 9.2 (repeated here from Chapter 7), the change from formal to semantic is transitional and especially slow in the case of hybrids. As I said in Chapter 7, this process is related to the understanding of socio-cultural realities. That is, how soon socio-cultural sex associations will be reflected in child grammar depends on the level of children’s cognitive development and their awareness of the importance of sex distinctions in gender assignment. I also proposed that two- and three-year-olds lack this socio-cultural knowledge, which explains the delay in the acquisition of the semantic rule for nouns in variable contexts. In Section 9.3.3, where I explain the differences in the acquisition of the semantic rule across the noun classes, I present additional considerations with regard to this delay.

Figure 9.2: Semantic agreement between younger children (25 children, age 2;6-4;0), older children (12 children, age 5;1-6;5), and adults (21 speakers): overall results for hybrids referring to females and female names in *-ok/-ik*



To conclude, the developmental sequence discussed in this section demonstrates that children do not approach the acquisition task with an intention to match the target grammar. They appear to be on the look-out for certain structural regularities. Their deviation from adult language and adult behavior in identical situations is thus not an imperfection but, as suggested by Yang (2000:234), “reflects principled hypotheses of language that children entertain before conclusively settling on the target language”. In the next section I propose that gender acquisition task, and the acquisition of the semantic rule in particular, consists of finding the appropriate cue in the primary linguistic data that children are exposed to.

9.3.2 Cue-based gender acquisition

The claim that children proceed from formal to non-formal analysis of the gender category leads to the following assumptions about the course of acquisition. Given that children base their initial hypothesis about gender on the formal properties of nouns, second declension masculine nouns (i.e. *papa*-type nouns, male names in *-a*, and double gender nouns) should be assigned feminine, while first declension nouns (i.e. hybrids and female names in *-ok/-ik*) should be assigned masculine. This means that at the earliest developmental stage (i.e. around age 2;0) children attribute nouns like e.g.

papa ‘daddy’ and *mama* ‘mommy’ to the same gender, since they are equivalent morphologically. In order to master the grammatical distinction between *papa* ‘daddy’ and *mama* ‘mommy’, children must realize that sex distinctions have grammatical realization as well. The important question is how children deduce that semantics plays a role in gender assignment? What triggers children to recognize the dominance of the semantic rules?

I suggest that in the process of acquisition children rely on certain cues which are derived from the input and reveal how the gender system is organized. Within the cue-based approach to language acquisition and change, cues have been defined by Lightfoot (1999:149) as “[...] abstract structures manifested in mental representations which result from parsing utterances”. More specifically, target input forms are used by language-learners as sources of cues which are not sets of sentences but mental representations in the internal grammar, I-language, which result from parsing utterances. With regard to gender, I propose that the child, who scans the linguistic environment and seeks for designated cues, first discovers the following:¹

- (1) [N-Ø V-Ø] and [N-*a* V-*a*]

These cues trigger the acquisition of morphological rules, i.e. they reveal the correspondence between the morphological properties of a noun and the morphological properties of a verb (or any other target). In other words, exposure to a phrase *stul upal* ‘the chair fell’ or *mama upala* ‘mommy fell’ should trigger the analysis of a noun as a controller. These cues trigger the acquisition of two distinct grammatical classes based on the specific morphological properties. They do not reveal anything about semantics, as there is a massive overlap of inanimate and animate nouns that have these representations. Therefore, only when the child finds the following cues, s/he can deduce that some other factor is at work here:

- (2) [N-Ø V-*a*] and [N-*a* V-Ø]

For the child who already knows that nouns like e.g. *stul* ‘chair’ or *mama* ‘mommy’ command agreement on adjacent (or non-adjacent) elements, exposure to phrases like *vrač prišla* ‘the physician came’ or *papa prišel* ‘daddy came’ should trigger the realization of biological sex as another factor that has grammatical function. As the mismatch between the cues in (1) and (2) reveals that the noun no longer controls the agreement, namely that formal

¹The cues in (1) represent verbal agreement. The corresponding cues for adjectival agreement may be stated as follows: [A-*ij/yj/oj* N-*a*] and [A-*aja/ija* N-*a*]. In what follows I will only consider cues for verbal agreement, which are sufficient to illustrate the problem under scrutiny.

properties of a noun do not correspond to those on the targets, the child should figure out what it is that commands the agreement. I suggest this is the point when the child can retrieve that semantic/sex distinctions also co-occur with already familiar inflectional morphemes.

The structures in (2) suggest that there should be more than one semantic cue. Likewise in other grammatical domains, e.g. verb second (V2) syntax, it has been argued for the existence of separate input cues and not a single global cue (cf. Westergaard (2006)).² Furthermore, I propose that each subtype of nouns should be represented by a separate semantic cue, such as those formulated in (3).³ The structures in (3) demonstrate that very specific information is included in the formal representation of each cue. In addition to the formal and semantic properties of a certain subcategory of nouns, the representations in (3) indicate whether gender is an inherent property of a noun or not. I use square brackets for nouns that have biological sex as part of their lexical semantic content. Round brackets indicate that gender is assigned via identification with a human referent.

- (3) a. $[_{[+male]}N-a V-\emptyset]$ (for *papa*-type nouns, male names in *-a*)
 b. $[_{(+male)}N-a V-\emptyset]$ (for double gender nouns referring to males)
 c. $[_{(-male)}N-\emptyset V-a]$ (for hybrids referring to females)
 d. $[_{[-male]}N-\emptyset V-a]$ (for female names in *-ok/-ik*)

With regard to the nouns under investigation, a cue-based learner should further distinguish the following cues, which express the opposite gender value for the same lexical form:

- (4) a. $[_{(-male)}N-a V-a]$ (for double gender nouns referring to females)
 b. $[_{[-male]}N-\emptyset V-\emptyset]$ (for female names in *-ok/-ik*)
 c. $[_{(\pm male)}N-\emptyset V-\emptyset]$ (for hybrids referring to males or females)

The semantic gender cues are thus manifested in the input through agreement and express evidence for different subtypes of nouns. In order to detect

²Westergaard (2006) extended the cue-based model proposed by Lightfoot (1999) to the acquisition of V2 vs. non-V2 syntax in different clause types in Norwegian (Tromsø dialect). She argues that there is no global cue for V2 syntax, as in the Lightfoot's model, instead each clause is represented separately. Furthermore, Westergaard (2006; forthcoming) argues that children have knowledge of the clause types and the corresponding syntactic heads (this information is provided by UG) but further information related to the position of the verb in a certain structure must be triggered by primary linguistic data in the forms of cues. Given this, with regard to gender it may be suggested that UG provides children not simply with the category N, but the knowledge that N can be of different subtypes. However, what these subtypes are must be learnt from the input.

³Binary feature representation is used here for convenience.

these cues, children should pay attention to the relevant subtype separately, to acquire its specific gender characteristics. As suggested by the course of acquisition, and specifically by the non-simultaneous acquisition of the semantic rule across the subtypes of nouns, children indeed look for specific cues and pay attention to the notion of class.

Interestingly, at some later point of development, sex distinctions become of such great importance for children that they start applying the semantic rule to inanimate nouns. Consider for example the following conversation of a four-year-old boy with his mommy where he refuses to use the feminine noun *carapina* ‘scratch’ to refer to himself; by changing the form of the noun to *carap* he intends to emphasize his sex - now, when the word ‘looks like’ a first declension masculine noun, he can use it about himself (from Čukovskij 1965:374):

- (5) a. CHILD: Mama, u menja na pal'ce **carap!**
 mom at I_{GEN} on finger_{LOC} scratch_{NOM}
 ‘Mom, I have a scratch on my finger!’
- b. MOT: Ne **carap**, a carapina.
 not scratch but scratch_{NOM(F)}
- c. CHILD: Eto u Musi esli, - carapina, a ja
 this at Musja_{GEN(F)} if scratch_{NOM(F)} but I_{NOM}
 mal'čik! U menja **carap!**
 boy_{NOM(M)} at I_{GEN} scratch
 ‘If it were about Musja, she would have *carapina*, but I am a boy!
 I have *carap*.’

Of course, such examples when children try to “regularize” the input are rare, nevertheless, they are straightforward in showing how with age, children who have realized the relevance of the semantic criterion for the language, make an attempt to reanalyze the grammatical gender system on semantic grounds.⁴

The semantic gender cues formulated in (3) and (4), and specifically, the structures in (3-a) and (3-b), appear to be formal representations of semantic rule proposed in Section 6.4.4, example (5). The reader may recall the proposal I made in Chapter 6, namely that the semantic rule formulated by Corbett (1991), e.g. ‘nouns denoting males are masculine’, may to be too

⁴The mismatches discussed above can be one of the sources for this analysis. Another source may be agreement with personal pronouns *ja* ‘I’ and *ty* ‘you’ which is only controlled by the biological sex of the speaker or the addressee. In other words, the child’s triggering experience for semantic gender distinctions can be the structures without N and where sex is the only possible controller.

general. A child, who entertains two competing hypotheses at the same time, might need a more specific semantic rule, such as ‘nouns of declensional class II denoting males are masculine’, that could override a more general morphological rule. The cues proposed in this section are thus formal representations of very specific semantic rules. The empirical facts about the acquisition of the semantic rule for various subcategories of nouns, and most importantly the differences in the course of acquisition, lend support for the idea of separate semantic cues for each subcategory of nouns. In the next section I discuss these important differences within the cue-based approach, and in particular, consider the role of input in this process.

9.3.3 Non-simultaneous acquisition of the semantic criterion

The most striking and obvious result displayed in Figure 9.1, which supports the idea that children distinguish classes of categories, is the contrastive use of semantic agreement in obligatory vs. variable contexts. In obligatory contexts children’s production of semantic agreement amounts to roughly 90% across the noun types, while in variable contexts it is below 20%. This means that having established the categorical status of the semantic rule for nouns in obligatory contexts, children do not immediately generalize it to other environments. Thus the acquisition of the semantic rule is not simultaneous for hybrids and female names in *-ok/-ik* on the one hand, and for masculine nouns in declension II and double gender nouns on the other. Furthermore, the prediction, discussed in Section 4.5, also suggested that the acquisition of the semantic rule would be delayed for hybrids, female names and double gender nouns. Clearly, it is borne out for hybrids and female names, but not for double gender nouns for which the semantic rules seem to be acquired during the considered developmental period, i.e. between 2;6-4;0. The differences in the acquisition of the semantic rule have also been noticed between various sub-types of nouns. Specifically, as shown in Chapter 6 and repeated in Figure 9.2 in this chapter, the acquisition of the semantic rule may be delayed for some low-frequency male kinship terms, but not for the low-frequency male names in *-a*. In addition, as demonstrated in Chapter 7, the semantic rule gains dominance faster for female names in *-ok/-ik* than for hybrids referring to females. Noteworthy, the semantic rule is not acquired simultaneously for proper names; its acquisition is delayed for the subclass of female names in *-ok/-ik* compared to the subclass of male names in *-a*.

How can the differences in the acquisition of the semantic rule be explained? More specifically, how can we account for the delay with regard

to hybrids and female names and explain the differences between these two? Why is there no delay for double gender nouns? And why should not all proper names be easy? Within the proposed cue-based approach to gender acquisition the answer to this question should be in the nature of the triggering experience. In the same way as in other studies of the delays in the acquisition of various grammatical phenomena, e.g. Westergaard's (forthcoming) study of subject shift in Norwegian child language, the factors which must be taken into account when answering this question are complexity of the pattern, the input frequency of the sentence types that express the specific cue, and the quality of the relevant input information.

Complexity

The complexity we should be looking at here is related to the following question: With regard to which noun classes is it most difficult to detect that a semantic rule plays a role? It is clearly not easy when the meaning of the noun does not provide information about the sex of the individual it refers to. In other words, when the same noun can be used to refer to the individuals of either sex. Given this, hybrids and double gender nouns must be more complex than *papa*-type nouns and male and female names. Nevertheless, a delay in the acquisition of the semantic rule has been found for hybrids but not for double gender nouns. It has also been found for female names in *-ok/-ik*, which, from this perspective, must be less complex than hybrids and double gender nouns, as they refer to females only. Yet, on the one hand, the acquisition of the semantic rule for this subtype of nouns is delayed compared to *papa*-type nouns and male names in *-a*, but on the other hand it seems to be acquired earlier than for hybrids. Thus, complexity does not seem to be a factor that could account for differences in the acquisition of the semantic rule across the subtypes of nouns. Therefore, in what follows I discuss other characteristics related to the quantitative as well as qualitative properties on the input, from which the specific cues are derived.

Frequency

Another factor that could potentially explain the differences is the input frequency of the sentence types that express the specific cue. My main concern here is whether the semantic cue is expressed on every occasion when the noun is used. The relevant cues formulated in the previous section are repeated in (6) for convenience.

- (6) a. $[\text{+male}]N\text{-}a\text{ V-}\emptyset$ (for *papa*-type nouns, male names in *-a*)

- b. $[(+male)N-a V-\emptyset]$ (for double gender nouns referring to males)
- c. $[(-male)N-\emptyset V-a]$ (for hybrids referring to females)
- d. $[(-male)N-\emptyset V-a]$ (for female names in *-ok/-ik*)

The semantic cue in (6-a) is certainly always evident in the input when *papa*-type nouns are used with some agreement target as well as male names in *-a*, since semantic agreement is obligatory in these cases. It can also be argued that these types of nouns are very frequent in child directed speech. The next cue in (6-b) triggers the acquisition of the semantic rule for double gender nouns, but I will postpone their discussion till later in this subsection. Consider now that in contrast to *papa*-type nouns and male names in *-a*, the semantic cue in (6-c) is not expressed on every occasion when a hybrid noun is used. It should refer to a female, but even this does not guarantee that the semantic cue will be evident, since semantic agreement is optional. The same is true for female names *-ok/-ik*, for which semantic agreement is optional in the input.⁵

Based on these observations it seems reasonable to conclude that frequency could play a role. But how do double gender nouns fit into this picture? Even if one assumes that as a type of nouns, double gender nouns are more frequent than hybrids or female names in *-ok/-ik*, so that the structures representing the semantic cue for double gender nouns could surface more frequently than for the other two types, the semantic cue in (6-b) may still not be expressed on each occasion when a double gender noun is used. Specifically, for the cue in (6-b) to be expressed, the noun should apply to a male and this may not be 100% of the time, as it is the case for *papa*-type nouns and male names in *-a*.⁶ Thus, given that the structures that express the semantic cue for double gender nouns are presumably less frequent than

⁵As a noun class female names in *-ok/-ik* may also be less frequent than e.g. male names in *-a*. Even among other derivatives from female names, forms in *-ok/-ik* seem to be less frequent options than forms ending in *-ka/-ša*. A Google search for diminutive forms for a female name *Marina* resulted in 117 000 for *Marinoč-ka* vs. 12 000 for *Marinč-ik*. From some names it is rather odd to derive a form in *-ok* or *-ik*, e.g. from a common name *Tanya*: ??*Tanik*, ??*Tan'čik*.

⁶It can be assumed that a boy is mostly exposed to masculine and a girl to feminine, since caregivers may mostly use double gender nouns to characterize their own children. Then it may be natural for boys to be biased towards masculine and for girls towards feminine. Some indication of this has been found in Chapter 8, but only for boys. Girls' behavior did not reveal a considerable overuse of feminine. As shown in Table 8.3 Chapter 8 they produced 90.4% masculine responses and 96.4% feminine responses. Additionally, in the task where the referent of a double gender noun was unknown, girls' responses were random. Therefore, to consider this assumption seriously, a more detailed investigation of both child and caregiver speech should be done. The presence of siblings of different sexes should also be taken into account.

for *papa*-type nouns and male names in *-a*, the frequency account should fail.

The notion of frequency has been reviewed in Chapter 4, where I mentioned that some researchers consider it an important factor, while others reject its importance for the acquisition process. In this study I have shown that token frequency of individual lexical items can affect the acquisition of the semantic rule with one class of common nouns (i.e. *papa*-type nouns), yet it has been argued to have no effect on the acquisition of proper names (i.e. male names in *-a*). Thus the role of frequency is not very clear, but it certainly cannot be rejected without consideration. Many researchers have come to the conclusion that frequency can only play a role in combination with other factors. For example, Kupisch (2007) and Westergaard and Bentzen (2007) argued that morphological or structural complexity is closely linked to frequency, and Roeper (2007) has explicitly shown that frequency account should be considered with caution. Therefore, in the next subsection I explore qualitative properties of the input and argue that consistency in the representation of the semantic cues is crucial and can account for the non-simultaneous acquisition of the semantic rule across the noun classes. Yet, in the absence of consistency the amount of input is rather important.

Qualitative properties of the input

Qualitatively the structures that represent the semantic cues in the input are different: in obligatory contexts the child's triggering experience is consistent in showing that only semantic agreement is available for *papa*-type nouns, male names in *-a*, and double gender nouns. In variable contexts, on the other hand, agreement forms are subject to variation, they have different frequencies of occurrence across individual nouns, speakers, style, as well as agreement target. This means even with reference to a female both masculine and feminine forms may occur in adult speech. Hence the child who is on the lookout for the semantic cues may find such information indeterminate. In other words, an inconsistent agreement pattern, unlike a consistent agreement pattern, can fail to provide sufficient evidence for the use of the semantic rule with hybrids referring to females and female names in *-ok/-ik*. Therefore, in the absence of consistency, the child must have extensive experience before s/he can extend the acquired knowledge of the semantic rule to a new domain.

Consistent evidence for the semantic cue $[[+male]N-a V-\emptyset]$ (cf. (6-a) above) appears to be sufficient for the establishment of the semantic rule at a relatively early age, as the highly adult-like usage of the semantic rule is evident for high-frequency *papa*-type nouns and male names in *-a* already at the age of 2;6. In addition, consistency of the triggering experience guar-

antees predominant use of the semantic rule for double gender nouns, where masculine and feminine forms are found in complementary distribution (masculine for a male and feminine for a female).

However, children do not seem to define the semantic rule in categorical terms: all sex-differentiable nouns must fulfill the semantic criterion. They appear to be conservative in their use of the semantic rule for hybrids referring to females and female names in *-ok/-ik*: at an early age they use it optionally and rather infrequently. The reason for this must be an insufficient amount of evidence in the child's primary linguistic data. Along the lines of Roeper (2007), who proposed that children use "incremental knowledge acquiring a list of contexts", one could argue that to become a more confident user of the semantic rule the child must receive ample evidence. Therefore, initially the children prefer to rely on morphology until enough information arises to decide which rule should play a dominant role. Importantly, there is no evidence in the data that children prioritize morphology with double gender nouns. Even when the referent's sex for these nouns was unspecified, the children in this study, and especially girls, were equally prone to use both masculine and feminine, while for hybrids they were prone to use only masculine. This observation together with the finding mentioned above (regarding the complementary relationship of masculine and feminine forms with double gender nouns) can be taken to indicate that children make a distinction between hybrids and double gender nouns.

Finally, it should be noted that the argument for consistency may also apply to female names in *-ok/-ik*, where the semantic rule appears to be acquired faster than for hybrids, as demonstrated in Figure 9.2. Recall from Chapter 7, that caregivers' production of the semantic (feminine) agreement with these nouns was near-categorical, i.e. 97.8%, while for hybrids referring to females it was used at a lower rate of 78.5%. This indicates that the semantic cue may be represented with greater consistency for female names than for hybrids, which should also be connected to the fact that the former apply only to females. Hence in the case of female names the evidence sufficient for the establishment of the dominant role of the semantic rule may be obtained earlier by children than in the case of hybrids. More evidence is needed to confirm this suggestion.

The conclusion I would like to make here is the following: in a language where the gender system is organized the way that the semantic rule is only crucial for very specific classes of nouns, the children appear to be conservative learners who use input information to make specific inferences concerning the gender of the nouns in their language and generalize only within a particular class.

The findings discussed so far lead to the reconsideration of the gender as-

signment theory. More specifically, they cast doubt on the idea of a semantic core as concerns the Russian gender system. In the next section I bring up learnability issues and argue against the core function of the semantic criterion.

9.4 Gender assignment and gender learnability

The developmental sequence discussed in Section 9.3.1 suggests that the children do not follow the computation which the gender assignment theory assumes for a language like Russian. That is, instead of prioritizing semantics from the start, they prioritize morphology. In Chapter 2 the Russian gender system was presented along the lines of Corbett (1991), who has given the most comprehensive account of gender systems. Recall that Corbett's point of departure is that "[...] all gender systems are semantic in that there is always a semantic core to the assignment system" (Corbett 1991:8).⁷ In a later work, Corbett and Fraser (2000) specify the distinction between semantic and formal systems as follows. They argue that "[t]he major distinction is between semantic systems (where only semantic information is required) and formal systems (where semantic information is supplemented by morphological and/or phonological information)" (Corbett and Fraser 2000:294). Hence in Russian, semantics is placed on top of the noun classification system, and morphology, which determines gender of the majority of nouns, plays the role of a supplement, i.e. it is secondary to the semantics. This may sound illogical, nevertheless, Russian is a good example of why this may be so: despite the fact that morphology allows correct gender assignment for most nouns and semantics for very few, it would not be possible to account for the gender of nouns like *papa* 'daddy', unless the dominance of semantics is admitted. Thus, the system, which is justified by a few exceptions, suggests that a learner should first try to establish a noun's gender on the basis of its semantics, and if this is not possible, s/he should make use of the noun's morphological characteristics. In other words, a learner should first decide whether a noun refers to a human being and, if yes, whether the referent is a male or a female. This allows the learner to assign either masculine or feminine gender to this noun. But if the noun does not refer to a human being, the learner should establish which declensional class it belongs to in order to infer its gender.

⁷This includes a number of possibilities, such as 'male/female', 'animate', 'insect' and so on.

Although Corbett does not make specific claims with regard to how the phenomenon of gender could be acquired, he assumes that “[w]e would expect semantic assignment systems to be acquired relatively easy” (Corbett 1991:82). In other words, he predicts that semantic gender systems should be acquired with ease as compared to the complex systems where formal and semantic criteria are interrelated. Given that semantic principles supersede, it can also be predicted that they should be acquired before formal principles, in order for children to avoid a forced retreat. Nevertheless, neither of these predictions seems to be borne out. As follows from the discussion of the previous language acquisition research considered in Chapter 3, not only is the acquisition of the strictly semantic (natural gender) systems (e.g. English) problematic, the acquisition of the target-consistent pronominal agreement in such languages is surprisingly late as compared to the same structures in languages with ‘mixed’ systems. Recall, for example, that in Mills’s (1986) study German monolinguals were in advance of English children of the same age in their production of third person singular pronouns. De Houwer (1990) also showed that in the speech of a Dutch-English bilingual child, Dutch pronouns were semantically determined before the English ones. Moreover, to the best of my knowledge there have not been found data showing that semantic (sex) criterion is acquired prior to the formal one, while there is a large amount of data from various languages showing that the acquisition proceeds from formal to semantic, as has also been proposed in this study.

The predictions do not get support from the present study either. If semantics is the core, children should be able to assign gender to the nouns in this study without difficulty simply by paying attention to the sex of their referents. If children used this straightforward procedure, they would not need to make false assumptions and retreat from them later. Nevertheless, the data in this study suggest that the child applies the logic which is the opposite from what the gender assignment theory predicts and which, as I propose below, seems to be innately specified. As we have seen in the previous chapters, children seem to realize that morphology is a powerful factor. Even in the case of *papa*-type nouns, which, unlike other nouns, denote only males and take consistent agreement, children assume that morphology is important and entertain this hypothesis for some time. As I have shown in Chapter 6, at the age of 2;6 children recognize the categorical status of the semantic rule for the high-frequency nouns, and yet overregularization errors still occur after this age for some low-frequency nouns. In addition, the findings reported in Chapter 7 suggest that when children are not sure about the status of the semantic rule, they prefer to rely on morphology for an extended period of time, as they do for hybrids referring to females and female names in *-ok/-ik*.

Corbett’s claim that semantics has to be a universal core is based on

typological evidence from languages with predominantly formal assignment systems, such as Russian, French, Swahili inter alia, where semantics is only crucial in some exceptional cases. For example, semantics is required to account for the masculine of Russian *papa* ‘daddy’, the feminine of French *femme* ‘woman’ or gender 1 of Swahili *kifaru* ‘rhinoceros’ (cf. Corbett 1991).⁸ Formal principles are still able to assign gender to the large proportion of nouns in these languages including those that denote humans and, in fact, in the absence of nouns like *papa* ‘daddy’, i.e. in the absence of form-meaning mismatches, the semantic rule would not be needed. The dominating role of semantics is thus restricted to the cases where morphology and/or phonology fail. It is rather surprising, then, that the latter are merely assigned a role of a supplement in Corbett’s account. In fact child data raise even more doubts. Admittedly, the knowledge of morphology is not sufficient in order to assign gender correctly to the nouns examined in this thesis, and children come to this conclusion in the course of acquisition. Nevertheless, the data from Russian and other languages suggest that morphology is not merely a supplement to semantics, but a very powerful factor from the first language-learner’s point of view. Consequently, if languages did not permit exceptions, how could children learn semantic rules? Morphology would be sufficient on its own to account for assignment and children would not even realize that semantics can play a role. Clearly, in the absence of such exceptions the semantic rules would be absolutely redundant and there would be no way the child could acquire them. Thus, from the point of learnability, semantics is unlikely to be the core.

An important observation related to the crucial role of formal criterion for gender assignment systems has been done by Polinsky (in press). Based on the evidence from heritage languages like e.g. American Russian, she predicts that “[i]n the absence of declensional information, one may expect changes in the assignment of nouns to gender classes” (Polinsky in press:28). One of such changes, according to her, is disappearance of gender as a grammatical category, which, for example, happens in creoles. Polinsky (1995) argues that a similar process takes place in heritage languages, where the knowledge of declension is either entirely absent or seriously reduced in heritage speakers.

Finally, Corbett’s theory runs into problems when one wants to account for the acquisition of generics like *rebenok* ‘child’ (masculine) or *persona* ‘person’ (feminine). These nouns denote humans and can refer to either male

⁸Russian nouns like *papa* ‘daddy’ belong to declensional class typical of feminines, but they are of masculine gender. Similarly, French nouns ending in [m] are predominantly masculine, but *femme* [fam] denotes a woman, hence it is feminine. Finally, in Swahili *kifaru* ‘rhinoceros’ ought to belong to the 7/8 gender due to its prefix *ki-*. However, it belongs to gender 1/2 as it is animate (cf. Corbett 1991).

or female individuals (just like the double gender noun *sirota* ‘orphan’), yet they always have only one gender, masculine or feminine, due to their morphological properties. This means that even when *rebenok* ‘child’ (masculine) is used non-generically to refer to a specific girl, all agreement forms (except personal pronouns) will be masculine. If the child is supposed to prioritize semantics, it is not clear how she can come to the understanding that it is morphology which is crucial here. In other words, in accordance with the semantic rules the child should assign the noun *rebenok* either masculine or feminine gender depending on the referent. But having learnt the gender of this noun in this way, how can a child reach the target state where its gender is always derived on formal grounds?

The acquisition data undermines the idea of a semantic core: formal criteria appear to be the governing factor rather than a supplement to semantics. The reasons why this is so are discussed in the next section, where I suggest that the process of gender acquisition is impossible without an innately organized circuitry. More specifically, I develop the idea that language acquisition is guided by innate abilities and mechanisms, but that input is necessary to infer language-specific characteristics.

9.5 Innateness and learning

Throughout this thesis I have been saying that the task of gender acquisition is complicated by the existence of input which is contradictory in terms of the cues it provides. But obviously experience plays an important role in gender acquisition, since the child must scan the input, searching for the relevant cues, and on the basis of this language-specific information s/he must formulate the specific rules. Furthermore, a learner should understand the relevant dimensions of meaning, e.g. natural gender, by using information from events and objects in the real world. Thus, gender requires inductive learning (Lust 2006) and may seem unlikely to be innate. Nevertheless, in this section I develop the idea that “[e]very bit of content is learned, but the system doing the learning works by a logic innately specified” Pinker (1999:210). More specifically, I propose that gender acquisition is an innately guided learning process (Gould and Marler 1987, Jusczyk 1997).

The idea behind innately guided learning proposed by Gould and Marler (1987) is that organisms are preprogrammed to learn particular things and to learn them in particular ways. According to them, honey bees, for example, are inherently suited to learning the odors, shapes and colors of different flowers. This means that information about flowers is in the genetic make-up of a bee, but the bee still has to learn which flowers are likely to hold food.

As soon as bees have acquired this knowledge about a flower, they organize and refer to that knowledge instinctively. Although bees have innate biases concerning which cues to rely on, such as odor, color and pattern, these cues are not remembered with equal weight. In fact, there is experimental evidence showing that bees' memory has hierarchical structure: odor takes precedence over color, and color takes precedence over shape. The structure of the bees' memory allows Gould and Marler (1987:197-198) to conclude that "instinct guides a bee's use of knowledge it gains through learning: [t]he cues that are memorized, the speed with which each cue is memorized and the way the memorized data are stored are all innate characteristics of a bee".

Likewise, gender acquisition can also be viewed as an innately controlled learning process. To start with I would like to go back to the finding that puzzled psycholinguists in the seventies and eighties, namely that children base their initial hypotheses about gender in various languages on formal, i.e. morphological and/or phonological criteria, while they disregard the semantic factor. This may be surprising, as from an adult perspective the concept of natural gender is simple and clear, while noun inflection seems to be a "memorizational nightmare" (Pinker 1999:213).

This finding questions the whole idea of semantic bootstrapping, which implies that language acquisition is based initially on children's ability to observe the world around them. Meaning is considered to be a critical factor, which forms the basis for formulating the linguistic characteristics of a particular phenomenon (cf. the discussion in Chapter 3). According to (Slobin 1973:187) "many linguistic forms cannot appear in the child's speech until he is capable of grasping their meaning". Apparently, gender is not one of these linguistic forms, since it appears to be rather easily accessible on formal grounds. From the start formal rules are used to assign gender, not only to inanimate nouns but also to nouns denoting humans, while language-external knowledge of the real world seems to be integrated somewhat later. Thus against all expectations the concept of natural gender is acquired as a supplement to the formal analyses. After all, this may not be surprising, since, as I pointed out in Section 9.4, in a language like Russian, the semantic information would not be needed if there were no words like *papa* 'daddy', etc. That is, children could easily determine a noun's gender from its declension type if there were no exceptions.

In other acquisition domains there is extensive evidence showing that rich agreement paradigms are acquired at a very early stage (Hyams 1986), that "... young children are extremely good at learning inflection" (Wexler 1998:43), and that "... children are sensitive to rich inflectional morphology [which gives them] a head-start for certain aspects of their linguistic development" (Baauw 2002:199). It is due to these striking abilities that they were

characterized as “little inflection machines” (Wexler 1998:43). Interestingly, these observations as well as the arguments presented in the previous section raise doubts against the view on morphology as a weak link expressed by McWhorter (2006). Upon this view inflectional affixation is a random epiphenomenon rather than a favored grammatical feature. Being an accident that entrenched in a language, it is seen as unnecessary. Yet, as follows from the discussion presented here, children acquiring their first language seem to share a different view on inflection.

With regard to gender, the question I would like to ask is the following: What makes a noun’s formal properties more easily accessible to the child than semantic properties? Several explanations have been mentioned in Chapter 3. In particular, the fact that young children attend to the noun’s form and use it predominantly to construct a powerful system of formal rules has been explained in terms of consistency (cf. Karmiloff-Smith 1979) and clarity of the rule (cf. Mills 1986). An alternative suggestion can be made in terms of Slobin’s (1973) operating principle A, which instructs a child to pay attention to the end of words while scanning the linguistic input for cues.⁹ These suggestions are plausible; yet, I would like to explore an alternative explanation.

Macnamara (1982) first noticed that children might be aware of the fact that the control of gender resides in nouns and not adjectives or articles. I further propose that knowledge that a noun carries an inherent gender specification and plays the role of gender controller, is innate. In other words, the fact that children first attend to the information expressed on the noun itself (and based on this make the appropriate gender generalizations very rapidly) suggests that this process may be innately guided.¹⁰ I now elaborate on the idea that children are innately predisposed to look for gender cues on the noun itself.

From Chomsky’s (1986) influential idea that children are innately equipped to distinguish between different grammatical categories, such as Noun (N), Verb (V), Adjective (Adj.), etc., it follows that among other lexical categories N must be innately recognized as the one that has a number of specific characteristics or Phi-features (Φ -features), i.e. gender, person, and number.

⁹Note here that principle A can only be relevant for languages like Russian where it is the ending of a noun that carries a gender marker. In Bantu languages, however, the gender marker is expressed by the prefix.

¹⁰It can be argued that semantic information must also come from the noun itself. Yet, biological sex seems to be a property of a human being, who happens to be a referent of a noun. In some cases it is not a property of a noun at all, e.g. professional titles or nouns like *crybaby*, but even for nouns that denote males or females like *daddy* or *mommy*, sex criterion is still realized by the association with a referent of a certain kind.

Since gender is not found universally, the acquisition task involves establishing which of the Φ -features are specified in the language the child is acquiring. In terms of Chomsky's principles and parameters approach (1981), children have to set the correct parameter based on information in the linguistic input. Since the information about grammatical gender is inherent in the linguistic make up of a noun, children first of all look for a gender defining feature on a noun itself. To put it differently, they are innately predisposed to look for gender cues on the nouns in the first place. This explains why in a linguistic situation related to the acquisition of grammatical gender, children rely on some cues (e.g. morphological or phonological) more readily than others (e.g. sex distinctions). Thus, it appears that innateness might be a stronger argument than consistency for why children base their initial hypotheses about gender on the formal characteristics of nouns.

However, as I said before, the process of gender acquisition also requires a certain amount of learning. Specifically, although children recognize formal cues innately, they still have to learn the inflections, since after all they are language specific. Thus, with respect to the formal gender specifications, it can be said that innately recognized cues trigger the learning process. On the other hand, the acquisition of the semantic concept of gender is purely cognitive learning. This type of knowledge cannot possibly be innate, because it comes from a nonlinguistic cognitive source; i.e. in order to derive information about the semantics of the words, children need to use information from objects and events in the real world. In the light of these crucial differences between the two gender criteria it seems to be clear why learning formal cues is easier and faster than learning the semantic cues: the former are grammar-internal, while the latter are extra-grammatical, and as such they have to be integrated into the language system. In other words, in the course of acquisition, children have to realize that extra-linguistic sex distinctions should be expressed by linguistic means. The results of the present study (similarly to previous observations) show that cognitive notions of gender are not accommodated into the grammatical system very early. According to the evidence in Chapter 7, the acquisition of the semantic rule is especially late when it is motivated by socio-cultural factors and does not have categorical status. This result again confirms that everything that is outside the grammar proper is taken into account later, since children's gender acquisition is guided by grammatical analysis from the start.

The differences between the two gender criteria also appear to be relevant crosslinguistically. Recall from Chapter 3 that complex morphological gender systems of languages like Polish, Russian, German, etc. are acquired rather easily, but not the semantic system of English, which is so simple from the second language learner's perspective. Yet, according to Mills (1986) and

De Houwer (1990), English three- and four-year-olds find it difficult, as the production of personal (*he/she*) and possessive (*her/his/him*) pronouns “appears to be full of errors” (Mills 1986:101). Surprisingly, German children are reported to be in advance in their production of masculine and feminine pronouns as compared to English children and, according to De Houwer, a Dutch-English bilingual child acquires semantically determined Dutch pronouns before the English ones.

To conclude, in this section gender acquisition has been characterized as an instance of innately guided learning: children are “preprogrammed” to learn particular gender characteristics in particular ways:

- In the course of gender acquisition the child subconsciously seeks for cues.
- Children are innately biased to treat N as gender controller, hence their early sensitivity to the formal (morphological and/or phonological) gender cues.
- There are two types of learning: grammatical and cognitive, which do not proceed in unison.
- Children inherently learn certain gender characteristics (i.e. formal) more easily than they learn the others (i.e. semantic properties related to natural gender, agreement).

Gender is thus an area of language which provides evidence that innateness and learning are not contradictory, but complementary in language acquisition.

Chapter 10

Summary and conclusions

10.1 Summary of individual chapters

In this dissertation I investigated children's ability to assign gender on the basis of the semantic criterion to certain exceptional classes of Russian nouns. Throughout this dissertation it became increasingly clear that the semantic rule is not acquired simultaneously for the noun classes under investigation, rather it seems to be acquired individually for every noun class. The detailed analysis of the experimental results revealed that various factors constrain the asymmetries in children's production for the individual noun classes. In a comprehensive analysis of the course of gender acquisition the differences are attributed to the qualitative and quantitative properties of the input.

The results of the experimentation are analyzed in three chapters: Chapter 6, Chapter 7, and Chapter 8. Each chapter aimed at answering specific research questions which were formulated separately in Chapter 4 and expanded upon at the beginning of each of the chapters of analysis.

Chapter 6 investigated children's behavior with masculine nouns in the second declension (*papa*-type nouns and male names in *-a*) and with the novel noun *obormoša* against the predictions of two theories of morphological acquisition: Words and Rules model (Pinker 1999) vs. Rules and Competition model (Yang 2002). I have shown that the acquisition of the semantic rule is the result of development, as consideration of semantic information increases with age. Although children in this study seem to have knowledge of the semantic rule already at an early stage (2;6-3;0), its full mastery for the subcategory of *papa*-type nouns can only be stated at approximately the age of 3;6. Token frequency in the sense of Pinker (1999) has been argued to be responsible for the asymmetry in the acquisition of the semantic rule across individual male kinship terms. The idea of 'free ride' effect in the

sense of Yang (2002) received no support from the data. I have also argued that the child's mechanism for gender acquisition is not limited to an initial item-based learning, as children's agreement production for the unfamiliar male names was target-consistent throughout the considered developmental period (2;6-4;0). Moreover, the children in this study could extend the acquired knowledge about their language to the novel nouns and invented words (as shown in Chapter 8). I proposed that frequency may not be the only factor responsible for the asymmetries in children's production. Children's early awareness of the differences in the nouns' semantic representation and specifically of the fact that proper names do not carry additional semantic content of the kind that common nouns have may account for the differences in the acquisition of male names in *-a* vs. *papa*-type nouns. Finally, my investigation in Chapter 6 allowed me to conclude that morphological and semantic gender principles might be qualitatively different. As follows from my experimental results, the integration of the semantic criterion proceeds gradually and involves some rote-based learning. However, gender assignment on the basis of morphology, as shown in previous acquisition studies of Russian and other languages, proceeds in a rule-based fashion.

Chapter 7 focused on the acquisition of the semantic rule for nouns where grammar allows certain variability. It was shown that between the ages of 2;6 and 4;0 the semantic rule is operative but not dominant for both hybrids referring to females and female names in *-ok/-ik*. Children's behavior in variable contexts is thus different from their behavior in obligatory contexts. This result was attributed to the lack of knowledge of socio-cultural constraints on use. Specifically, children seem to lack the ability to use the semantic rule to produce social meaning, while they are able to use it for grammatical purposes, i.e. when the grammar requires it. This result also suggested that children may be sensitive to the qualitative/quantitative differences in the input for individual subtypes of nouns. The implications of this view on the child's data were discussed in Chapter 9. The data also revealed that children's behavior is different from their caregiver's who used the semantic rule predominantly for hybrids referring to females and near-categorically for female names in *-ok/-ik* in the same experimental situation. This result was interpreted along the lines of Kerswill and Williams (2000) among others, namely that children need to mature socially in order to be able to use the semantic rule for these subcategories of nouns. Furthermore, I proposed that for two- and three-year-olds gender might be a purely grammatical category and not a socio-linguistic phenomenon. Finally, I made a comparison of the agreement production from younger (2;6-4;0) and older children (5;1-6;5) the results of which seem to support the hypothesized course of development from formal to semantic. In particular, in the speech of younger children I

found that the semantic rule was used optionally and less frequently than the morphological rule, which played a dominant role at this stage (2;6-4;0). At a later stage, between the ages of 5;1 and 6;5, it became evident that the semantic rule gains dominance gradually and very slowly for these subtypes of nouns. In addition, I found that the developmental curves for female names vs. hybrids were different, as the semantic rule gained dominance faster for female names, while for hybrids the improvement was rather unnoticeable. This result was shown to be parallel to the result reported in Chapter 6: proper names in both obligatory and variable contexts behave differently from other subtypes of nouns; yet I suggested that these differences may not be accounted for in the same way.

Chapter 8 investigated children's knowledge of referential gender, i.e. children's ability to assign gender via identification with a specific referent on each occasion when the noun is used. The complementary relationship of masculine and feminine forms with the respective referent in children's production as well as their ability to assign two genders to the same noun depending on the biological sex of its referent was taken that two- and three-year-olds have knowledge of referential gender. Furthermore, I showed that the children in this study were not prone to use the morphological rule with double gender nouns even when the sex of their referents was not specified. Boys were found to behave differently from girls with this class of nouns, as they seemed to be predisposed to assign masculine based on their own sex, which was also found in adults. Yet, I suggested that there may be additional factors rather than the speaker's own sex that were involved here (e.g. inattentiveness). On the whole, children's production was still different from that of the adults. It was also different from their production for hybrids in the same experimental conditions. The results presented in this chapter were thus taken to indicate that children consider double gender nouns as a distinct class, where sex of the referent plays a dominant role and should be taken into account on every occasion when a nouns is used and where a noun can be assigned two genders, masculine or feminine.

Chapter 9 reconsidered the results of the experimentation to provide a more complete account of the gender acquisition process. First, based on the experimental results discussed in chapters 5, 6, and 7, I argued for a single underlying process, whereby formal grammatical analysis based on the formal gender properties is gradually replaced by non-formal analysis based on the semantic/sex-based properties. Second, I proposed that the acquisition of the semantic principle may be triggered by the abstract cues which are expressed in certain input structures. I furthermore suggested that for the child to discover the particular cue there should consistent evidence for it in the input. In the absence of consistency a child must have extensive experience

for the rule to be acquired. The idea of selective search for cues may thus explain the differences in the acquisition of semantic rule for various subtypes of nouns. One of the important claims that I made further in this chapter is related to the issue of learnability and, more specifically, the core function of the semantic criterion in a gender system of Russian. The evidence from the acquisition of gender in Russian has been shown to undermine the idea of a semantic core, since formal criteria appear to be crucial for the establishment of the gender system in a language and thus unlikely to be a mere supplement to semantics. Finally, I explored the idea of gender acquisition as an innately guided learning process, in the course of which children are “preprogrammed” to learn particular things and learn them in particular ways.

10.2 Ideas for further research

Based on the evidence obtained in Chapters 6 and 7, I suggest that the role of input in discovering the gender cues should be investigated further, in particular with regard to the acquisition of proper names. I have found that the acquisition of gender with this subtype of nouns proceeds differently from other types, such as male kinship terms and professional titles. This raised additional questions, some of which are still open and some of the proposed suggestions need further empirical support.

In Chapter 6 I expressed the idea that the occurrence of specific strings in the input, such as ‘common noun+proper name’, may facilitate the acquisition of gender for the individual male kinship terms which occur in the adult speech in a combination with a proper name. Such combinations are presumably frequent in child-directed speech for the nouns *papa* ‘daddy’, *deduška* ‘grandad’, and *djadja* ‘uncle/man’, while the nouns *mužčina* ‘man’ and *junoša* ‘youth’ are unlikely to be used the same way. I suggest that this idea should be investigated further through comparing the child’s agreement production with these nouns to the input forms in the speech of a caregiver.

In Chapter 7 it was found that between the ages of 2;6 and 4;0 the children (in contrast to their caregivers) are not prone to use the semantic rule for hybrids and female names in *-ok/-ik*. Older children, between the ages of 5;1 and 6;5, were able to use the semantic rule to a greater extent with female names but not with hybrids. However, it remains unclear when and how the semantic rule becomes dominant for female names and especially for hybrids. Thus, it would be interesting to look at even older children (the children in their early school years). It is also unclear what the exact input with regard to these nouns is. In other words, in order to have a more clear picture of how variability in the input affects the acquisition of gender, it seems

necessary to compare the data from adults in spontaneous conversation to their production in the experimental situation.

In Chapter 8, based on the results of children's agreement production with double gender nouns I argued that children were using the semantic (male-female) distinction as a guideline in their choice of agreement forms for both male and female referents. However, it was pointed out that in the case of a female referent, when morphology and semantics overlap, one cannot be absolutely sure that the child uses the semantic and not the morphological rule. Therefore, it is possible that a comprehension technique may provide further details with regard to children's knowledge of the importance of the male-female distinction for referential gender. Specifically, I propose to explore the reverse, i.e. whether children can infer the sex of the individual by attending to masculine and feminine agreement forms.

Furthermore, it would also be interesting to compare the results of this study to the data from other languages, especially with regard to hybrids and double gender nouns. Of particular interest are of course languages with mixed gender assignment systems.

Finally, in order to have a complete picture of how the child acquires a gender system of Russian, one has to consider nouns ending in palatalized consonants, that can be either masculine or feminine, as well as neuter nouns with stressed and unstressed stems. In Chapters 2 and 3 it was mentioned that these nouns are problematic for a Russian child due to their ambiguous morphological form in nominative singular. According to Gvozdev (1961), their acquisition is delayed as compared to other masculine and feminine nouns, even those like *papa* 'daddy'. It would thus be interesting to account for the acquisition of gender with these nouns in the context of the exceptional classes of nouns considered in this study. These nouns also present interest regarding the question how children establish the relationship between declension and gender.

10.3 Concluding remarks

In this dissertation I have explored the relationship between formal and semantic criteria in the acquisition of the Russian gender system. Gender in Russian and gender systems in general have received considerable attention in theoretical linguistics and language typology. To my knowledge, this dissertation is the first large scale investigation of the interaction of formal and semantic criteria in the acquisition of Russian gender. This empirical investigation yields implications for the study of Russian, the study of gender in general, as well as language acquisition.

With regard to the gender system of Russian, this research touches upon a very important question of what the core of the system is. Counter to the general theoretical view, children seem to assume that morphology is the core and start building the system on formal grounds discovering morphological gender regularities in a very short time and overgeneralizing this knowledge to the domains where semantics should be dominant. The children thus behave as if semantics were secondary and they were forced to use it, but when there are no obligatory restrictions they revert to morphology. In the course of my investigation it also becomes clear that the semantic rules used by children may be more specific than what has been assumed. The rules seem to be formulated “locally” by children with regard to each particular noun class based on the relevant agreement information. It would thus be interesting to find out whether adults have similar local generalizations in their mental grammars, or whether their mental grammars are different from children’s grammars.

For the study of gender as a grammatical phenomenon this empirical investigation reveals that the interaction between semantic and formal criteria is much more intricate than one might think. Being initially built on formal grounds the system organization gradually changes from formal to semantic, and from lexical meaning to social factors. It would thus be interesting to test the problems related to the relationship of the gender criteria in other languages to see whether these results would be confirmed.

With regard to language acquisition, the research methodology adopted in this study allowed me to examine and compare certain exceptional classes of nouns which may be rare in spontaneous speech. The experimental approach made it possible to reveal that each subtype of nouns is acquired individually by children. This study thus supports the idea that children are sensitive to classes of categories. The results of the study provide novel support for the claim that the acquisition of gender is guided by grammatical analyses. Similar to the previous work on the acquisition of gender it shows that the system organization changes from formal to semantic. Most interestingly, it shows that children are sensitive to obligatory vs. variable contexts and that (natural) gender is first of all a grammatical category rather than a socio-linguistic phenomenon. Finally, the results of the study confirm that children are creative language learners. On the basis of input they receive, they formulate hypotheses, consider evidence, and make revisions. It has thus been proposed that the child approaches the gender acquisition task with the logic that may be innately specified.

Above all, this dissertation has shown that empirical investigations of the acquisition of the complexities of gender in Russian is a fruitful area of research with implications for a number of important issues. It is my hope

that this study will offer inspiration for further research on the acquisition of gender in Russian and other languages.

Appendix I

Table 10.1: The list of participants of the main study including their age and sex (25 children, age 2;6-4;0)

Child	Age	Sex
1. Dima	2;6	M
2. Olya	2;7	F
3. Petya	2;8	M
4. Roma	2;10	M
5. Katya	2;11	F
6. Seva	3;0	M
7. Lena	3;1	F
8. Kolya	3;1	M
9. Ira	3;2	F
10. Vasya	3;3	M
11. Lyuba	3;3	F
12. Nadya	3;3	F
13. Slava	3;6	M
14. Lera	3;6	F
15. Tolya	3;6	M
16. Vova	3;7	M
17. Inna	3;9	F
18. Galya	3;9	F
19. Vera	3;9	F
20. Liza	3;9	F
21. Denis	3;9	M
22. Polyu	3;10	F
23. Sonya	3;10	F
24. Roza	3;10	F
25. Oleg	4;0	M

Table 10.2: The list of older participants including their age and sex (12 children, age 5;1-6;5)

Child	Age	Sex
1. Borya	5;1	M
2. Valya	5;1	F
3. Vitya	5;3	M
4. Paša	5;3	M
5. Sveta	5;4	F
6. Nina	5;7	F
7. Anya	5;8	F
8. Stas	5;9	M
9. Jana	5;11	F
10. Yulya	5;11	F
11. Tanya	6;3	F
12. Dusya	6;5	F

Appendix II

Table 10.3: Agreement production with the novel noun *obormoša* (25 children, age 2;6-4;0)

Child	Age	ATTRIB ADJ		PST V		Prn		Total M/F
		M	F	M	F	M	F	
1. Dima	2;6	0	4	0	2	0	0	0/6
2. Olya	2;7	7	0	8	0	4	0	19/0
3. Petya	2;8	12	0	11	0	3	0	26/0
4. Roma	2;10	4	0	8	0	1	1	13/1
5. Katya	2;11	0	0	6	1	0	0	6/1
6. Seva	3;0	3	0	7	0	0	0	10/0
7. Lena	3;1	5	0	7	0	3	0	15/0
8. Kolya	3;1	6	3	0	9	0	0	6/12
9. Ira	3;2	4	0	10	0	2	0	16/0
10. Vasya	3;3	3	0	3	1	2	0	8/1
11. Lyuba	3;3	7	0	4	0	4	0	15/0
12. Nadya	3;3	13	0	6	0	0	0	19/0
13. Slava	3;6	4	1	11	0	3	0	18/1
14. Lera	3;6	5	0	7	0	3	1	15/1
15. Tolya	3;6	3	3	8	2	2	1	13/6
16. Vova	3;7	5	0	8	1	2	0	15/1
17. Inna	3;9	5	0	11	0	4	0	20/0
18. Galya	3;9	0	9	0	7	0	2	0/18
19. Vera	3;9	5	0	9	0	0	0	14/0
20. Liza	3;9	5	0	10	0	3	0	18/0
21. Denis	3;9	5	0	8	1	0	0	13/1
22. Polya	3;10	6	0	16	0	8	0	30/0
23. Sonya	3;10	4	0	11	0	3	0	18/0
24. Roza	3;10	4	0	8	0	3	0	15/0
25. Oleg	4;0	6	0	7	0	2	0	15/0
Total		184	24	121	20	52	5	357/49

Table 10.4: Agreement production with male names in *-a* (25 children, age 2;6-4;0)

Child	Age	M	F
1. Dima	2;6	2	2
2. Olya	2;7	6	0
3. Petya	2;8	6	0
4. Roma	2;10	10	0
5. Katya	2;11	11	0
6. Seva	3;0	12	0
7. Lena	3;1	6	1
8. Kolya	3;1	4	3
9. Ira	3;2	10	0
10. Vasya	3;3	4	1
11. Lyuba	3;3	10	0
12. Nadya	3;3	6	1
13. Slava	3;6	9	0
14. Lera	3;6	9	0
15. Tolya	3;6	9	0
16. Vova	3;7	12	0
17. Inna	3;9	10	0
18. Galya	3;9	9	1
19. Vera	3;9	10	0
20. Liza	3;9	13	0
21. Denis	3;9	11	0
22. Polya	3;10	11	0
23. Sonya	3;10	8	0
24. Roza	3;10	9	0
25. Oleg	4;0	10	0
Total		217	9

Table 10.5: Agreement production with masculine noun *papa* ‘daddy’ (25 children, age 2;6-4;0)

Child	Age	ATTRIB ADJ		PST V		Prn		Total M/F
		M	F	M	F	M	F	
1. Dima	2;6	6	0	0	0	0	0	6/0
2. Olya	2;7	4	0	0	0	0	0	4/0
3. Petya	2;8	5	0	2	0	0	0	7/0
4. Roma	2;10	5	0	0	0	0	0	5/0
5. Katya	2;11	3	1	0	0	0	0	3/1
6. Seva	3;0	1	0	0	0	0	0	1/0
7. Lena	3;1	3	0	0	0	0	0	3/0
8. Kolya	3;1	5	0	1	3	0	0	5/3
9. Ira	3;2	1	1	0	0	1	0	2/1
10. Vasya	3;3	4	1	0	0	0	0	4/1
11. Lyuba	3;3	5	0	0	0	0	0	5/0
12. Nadya	3;3	6	0	0	0	0	0	6/0
13. Slava	3;6	0	3	1	0	0	0	1/3
14. Lera	3;6	1	0	1	0	1	0	3/0
15. Tolya	3;6	5	0	2	0	1	0	8/0
16. Vova	3;7	6	0	1	0	0	0	7/0
17. Inna	3;9	8	0	2	0	1	0	11/0
18. Galya	3;9	4	0	0	0	0	0	4/0
19. Vera	3;9	6	0	1	0	2	0	9/0
20. Liza	3;9	5	0	1	0	0	0	6/0
21. Denis	3;9	5	0	1	0	0	0	6/0
22. Polya	3;10	7	0	0	0	0	0	7/0
23. Sonya	3;10	6	0	1	2	0	0	7/2
24. Roza	3;10	8	0	1	0	0	0	9/0
25. Oleg	4;0	1	0	0	0	0	0	1/0
Total		110	6	15	5	6	0	131/11

Table 10.6: Agreement production with masculine noun *djadja* ‘uncle/man’
(25 children, age 2;6-4;0)

Child	Age	ATTRIB ADJ		PST V		Prn		Total M/F
		M	F	M	F	M	F	
1. Dima	2;6	4	0	0	1	0	0	4/1
2. Olya	2;7	5	0	1	0	0	0	6/0
3. Petya	2;8	6	0	2	0	0	0	8/0
4. Roma	2;10	5	0	3	0	1	0	9/0
5. Katya	2;11	6	0	2	0	0	0	8/0
6. Seva	3;0	3	0	0	0	1	0	4/0
7. Lena	3;1	5	0	2	0	0	0	7/0
8. Kolya	3;1	4	1	3	2	0	2	7/5
9. Ira	3;2	5	0	4	0	0	0	9/0
10. Vasya	3;3	4	0	1	0	0	0	5/0
11. Lyuba	3;3	6	0	0	0	0	0	6/0
12. Nadya	3;3	6	0	1	0	0	0	7/0
13. Slava	3;6	2	1	3	0	0	0	5/1
14. Lera	3;6	4	0	0	1	0	0	4/1
15. Tolya	3;6	5	1	1	0	0	0	6/1
16. Vova	3;7	4	0	1	0	0	0	5/0
17. Inna	3;9	5	0	2	0	0	0	7/0
18. Galya	3;9	4	0	1	0	0	0	5/0
19. Vera	3;9	5	0	2	0	0	0	7/0
20. Liza	3;9	7	0	1	0	0	0	8/0
21. Denis	3;9	6	0	1	0	0	0	7/0
22. Polya	3;10	8	0	1	0	1	0	10/0
23. Sonya	3;10	8	0	2	0	0	0	10/0
24. Roza	3;10	5	1	2	0	0	0	7/1
25. Oleg	4;0	6	0	3	0	0	0	9/0
Total		128	4	39	4	3	2	170/10

Table 10.7: Agreement production with masculine noun *deduška* ‘grandad’ (25 children, age 2;6-4;0)

Child	Age	ATTRIB ADJ		PST V		Prn		Total M/F
		M	F	M	F	M	F	
1. Dima	2;6	3	0	0	0	0	0	3/0
2. Olya	2;7	4	0	1	0	0	0	5/0
3. Petya	2;8	6	0	1	0	0	0	7/0
4. Roma	2;10	6	0	1	0	0	0	7/0
5. Katya	2;11	1	0	1	0	0	0	2/0
6. Seva	3;0	2	0	2	1	0	0	4/1
7. Lena	3;1	4	0	0	3	0	0	4/3
8. Kolya	3;1	5	0	0	2	0	0	5/2
9. Ira	3;2	2	0	1	0	0	0	3/0
10. Vasya	3;3	6	0	1	0	1	0	8/0
11. Lyuba	3;3	4	0	0	0	0	0	4/0
12. Nadya	3;3	5	0	2	0	0	0	7/0
13. Slava	3;6	0	1	1	0	0	0	1/1
14. Lera	3;6	4	0	1	1	0	0	5/1
15. Tolya	3;6	3	0	1	0	0	0	5/0
16. Vova	3;7	2	0	0	0	0	0	2/0
17. Inna	3;9	2	0	1	0	3	0	6/0
18. Galya	3;9	4	0	1	0	0	0	5/0
19. Vera	3;9	3	2	1	0	0	0	4/2
20. Liza	3;9	4	0	0	0	0	0	4/0
21. Denis	3;9	1	0	2	0	0	0	3/0
22. Polya	3;10	6	0	0	0	0	0	6/0
23. Sonya	3;10	6	0	0	0	0	0	6/0
24. Roza	3;10	4	0	1	0	0	0	5/0
25. Oleg	4;0	6	0	4	0	0	0	10/0
Total		93	3	23	7	4	0	120/10

Table 10.8: Agreement production with masculine noun *mužčina* ‘man’ (25 children, age 2;6-4;0)

Child	Age	ATTRIB ADJ		PST V		Prn		Total M/F
		M	F	M	F	M	F	
1. Dima	2;6	1	3	1	3	0	0	2/6
2. Olya	2;7	4	0	2	0	0	0	6/0
3. Petya	2;8	5	0	0	0	0	0	5/0
4. Roma	2;10	3	0	0	0	0	0	3/0
5. Katya	2;11	3	0	2	0	0	0	5/0
6. Seva	3;0	1	1	0	0	0	0	1/1
7. Lena	3;1	5	0	3	0	0	0	8/0
8. Kolya	3;1	5	1	0	2	0	3	5/6
9. Ira	3;2	1	0	0	1	0	0	1/1
10. Vasya	3;3	4	0	1	0	3	0	8/0
11. Lyuba	3;3	4	0	0	0	1	0	5/0
12. Nadya	3;3	6	0	0	0	0	0	4/0
13. Slava	3;6	0	1	2	0	1	0	3/1
14. Lera	3;6	2	0	1	0	1	0	4/0
15. Tolya	3;6	2	1	4	0	0	0	6/1
16. Vova	3;7	2	0	1	0	0	0	3/0
17. Inna	3;9	6	0	1	0	1	0	8/0
18. Galya	3;9	3	0	1	0	0	0	4/0
19. Vera	3;9	3	0	0	0	0	0	3/0
20. Liza	3;9	3	0	2	0	1	0	6/0
21. Denis	3;9	5	0	1	0	0	0	6/0
22. Polya	3;10	12	0	3	0	3	0	18/0
23. Sonya	3;10	5	0	1	0	2	0	8/0
24. Roza	3;10	3	2	1	0	3	0	7/2
25. Oleg	4;0	5	0	1	0	0	0	6/0
Total		91	9	28	6	16	3	135/18

Table 10.9: Agreement production with masculine noun *junoša* ‘youth’ (25 children, age 2;6-4;0)

Child	Age	ATTRIB ADJ		PST V		Prn		Total M/F
		M	F	M	F	M	F	
1. Dima	2;6	3	2	0	2	0	0	3/4
2. Olya	2;7	4	0	1	0	0	0	5/0
3. Petya	2;8	5	0	1	0	0	0	6/0
4. Roma	2;10	1	0	2	0	0	0	3/0
5. Katya	2;11	3	0	1	0	0	0	4/0
6. Seva	3;0	2	0	1	0	0	0	3/0
7. Lena	3;1	4	0	1	0	0	0	5/0
8. Kolya	3;1	1	2	1	3	1	1	3/6
9. Ira	3;2	4	0	3	0	0	0	7/0
10. Vasya	3;3	5	0	0	0	1	0	6/0
11. Lyuba	3;3	4	0	0	0	0	0	4/0
12. Nadya	3;3	3	0	1	0	0	0	4/0
13. Slava	3;6	4	0	0	0	0	0	4/0
14. Lera	3;6	3	0	0	0	1	0	4/0
15. Tolya	3;6	3	0	2	0	0	0	5/0
16. Vova	3;7	7	0	1	0	0	0	8/0
17. Inna	3;9	4	0	1	0	3	0	8/0
18. Galya	3;9	3	0	1	0	0	0	4/0
19. Vera	3;9	4	0	0	0	0	0	4/0
20. Liza	3;9	5	0	2	0	0	0	7/0
21. Denis	3;9	2	0	1	0	0	0	3/0
22. Polya	3;10	4	0	2	0	2	0	8/0
23. Sonya	3;10	5	0	1	0	2	0	8/0
24. Roza	3;10	5	0	1	0	0	0	6/0
25. Oleg	4;0	4	0	0	0	0	0	4/0
Total		92	4	24	5	10	0	126/10

Table 10.10: Agreement production for female names in *-ok/-ik* (25 children, age 2;6-4;0)

Child	Age	M	F	Total
1. Dima	2;6	0	1	1
2. Olya	2;7	3	0	3
3. Petya	2;8	10	0	10
4. Roma	2;10	14	1	15
5. Katya	2;11	11	0	11
6. Seva	3;0	7	1	8
7. Lena	3;1	13	0	13
8. Kolya	3;1	2	6	8
9. Ira	3;2	12	0	12
10. Vasya	3;3	5	2	7
11. Lyuba	3;3	9	0	9
12. Nadya	3;3	11	1	12
13. Slava	3;6	6	0	6
14. Lera	3;6	9	0	9
15. Tolya	3;6	1	6	7
16. Vova	3;7	7	0	7
17. Inna	3;9	10	1	11
18. Galya	3;9	10	0	10
19. Vera	3;9	11	0	11
20. Liza	3;9	1	7	8
21. Denis	3;9	9	0	9
22. Polya	3;4	4	6	10
23. Sonya	3;10	9	0	9
24. Roza	3;10	11	0	11
25. Oleg	4;0	6	0	6
Total		191	32	223

Table 10.11: Caregiver's agreement production for female names in *-ok/-ik*.
(21 adult speakers)

Adult	M	F	Total
1. Dima MOT	0	6	6
2. Olya MOT	0	11	11
3. Roma MOT	0	8	8
4. Katya MOT	1	8	9
5. Seva MOT	0	8	8
6. Lena MOT	0	11	11
7. Kolya MOT	0	6	6
8. Ira MOT	0	8	8
9. Lyuba MOT	0	9	9
10. Nadya GR	3	7	10
11. Slava MOT	0	10	10
12. Lera MOT	0	8	8
13. Tolya GR	0	7	7
14. Vova MOT	0	11	11
15. Inna MOT	0	7	7
16. Galya MOT	0	8	8
17. Vera MOT	0	9	9
18. Liza MOT	0	7	7
19. Denis FAT	0	8	8
20. Polya GR	0	11	11
21. Sonya SIS	0	7	7
Total	4	175	179

Table 10.12: Agreement production with the female names in *-ok/-ik* (12 children, age 5;1-6;5)

Child	Age	M	F	Total
1. Borya	5;1	5	4	9
2. Valya	5;1	10	0	10
3. Vitya	5;3	0	5	5
4. Paša	5;3	7	2	9
5. Sveta	5;4	3	6	9
6. Nina	5;7	8	0	8
7. Anya Z.	5;8	0	9	9
8. Stas	5;9	2	7	9
9. Jana	5;11	0	8	8
10. Yulya S.	5;11	0	7	7
11. Tanya	6;3	0	7	7
12. Dusya	6;5	2	5	7
Total		37	60	97

Table 10.13: Agreement production with individual hybrid nouns referring to females: *počtal'on* 'postwoman', *doktor* 'doctor', *milicijoner* 'policewoman', *fotograf* 'photographer', *povar* 'cook', *vrač* 'physician', *bibliotekar'* 'librarian' (25 children, age 2;6-4;0)

Child	Age	<i>počtal'on</i>		<i>doktor</i>		<i>milicijoner</i>		<i>fotograf</i>		<i>povar</i>		<i>vrač</i>		<i>bibliotekar'</i>		Tot. M/F
		M	F	M	F	M	F	M	F	M	F	M	F	M	F	
1. Dima	2;6	0	0	0	2	0	3	0	0	1	0	0	1	0	1	1/7
2. Olya	2;7	1	0	1	0	2	0	0	0	2	0	0	1	3	0	9/1
3. Petya	2;8	2	0	2	0	2	0	0	0	2	0	0	1	1	1	9/2
4. Roma	2;10	1	1	3	0	4	0	1	0	4	0	2	0	0	0	15/1
5. Katya	2;11	1	0	1	0	2	0	1	0	2	0	0	2	0	2	7/4
6. Seva	3;0	1	0	2	0	0	0	0	0	0	0	0	1	2	1	5/2
7. Lena	3;1	2	0	3	0	4	0	2	0	2	0	1	0	0	0	14/0
8. Kolya	3;1	1	1	0	1	0	1	1	0	0	2	0	0	0	2	2/7
9. Ira	3;2	0	0	2	0	2	0	1	0	2	0	1	1	0	1	8/2
10. Vasya	3;3	1	0	1	0	3	0	1	0	0	0	2	0	1	0	9/0
11. Lyuba	3;3	1	1	2	0	2	0	1	0	1	0	2	0	1	1	10/2
12. Nadya	3;3	2	0	3	1	2	0	0	2	0	0	0	2	0	2	7/7
13. Slava	3;6	2	0	1	0	0	0	1	0	2	0	0	2	1	0	7/2
14. Lera	3;6	2	0	2	0	2	0	2	0	2	0	1	1	2	0	13/1
15. Tolya	3;6	0	0	0	0	0	1	0	1	2	0	0	2	2	0	4/4
16. Vova	3;7	2	0	3	0	3	0	1	0	2	0	0	3	1	1	12/4
17. Inna	3;9	2	0	2	0	2	0	1	0	2	0	2	0	2	0	13/0
18. Galya	3;9	3	0	3	0	2	0	1	0	2	0	0	2	2	0	13/2
19. Vera	3;9	2	0	1	0	2	0	2	0	2	0	2	0	2	1	13/1
20. Liza	3;9	2	0	2	1	2	0	2	0	3	0	0	3	1	1	12/5
21. Denis	3;9	2	0	3	0	2	0	1	0	2	0	2	0	1	0	13/0
22. Polya	3;10	2	0	1	0	2	0	1	0	2	0	3	0	2	0	13/0
23. Sonya	3;10	2	0	2	0	3	0	1	0	2	0	2	0	2	0	14/0
24. Roza	3;10	2	0	2	0	3	0	1	0	2	0	2	0	2	0	14/0
25. Oleg	4;0	2	0	2	0	2	0	1	0	2	0	0	1	0	2	9/3
Total		38	3	44	5	48	5	23	3	43	2	22	23	28	16	246/57

Table 10.14: Caregiver's agreement production with individual hybrid nouns referring to females: *počtal'on* 'postwoman', *doktor* 'doctor', *milicioner* 'policewoman', *fotograf* 'photographer', *povar* 'cook', *vrač* 'physician', *bibliotekar'* 'labrarian' (21 adult speakers)

Adult	<i>počtal'on</i>		<i>doktor</i>		<i>milicioner</i>		<i>fotograf</i>		<i>povar</i>		<i>vrač</i>		<i>bibliotekar'</i>		Total M/F
	M	F	M	F	M	F	M	F	M	F	M	F	M	F	
1. Dima MOT	0	2	3	0	0	2	0	0	2	0	1	1	0	2	6/7
2. Olya MOT	0	0	0	3	0	4	0	0	1	0	0	2	0	1	1/10
3. Roma MOT	0	1	0	3	0	2	0	1	0	0	0	2	0	1	0/10
4. Katya MOT	0	1	0	2	0	4	0	1	0	2	1	1	0	0	1/11
5. Seva MOT	0	1	0	3	2	0	2	0	2	0	0	2	0	0	6/6
6. Lena MOT	0	2	0	2	0	2	0	1	1	0	0	2	0	2	1/11
7. Kolya MOT	0	2	0	0	0	0	0	1	2	0	1	0	0	2	3/5
8. Ira MOT	0	1	0	1	0	2	0	1	0	1	0	2	0	1	0/9
9. Lyuba MOT	0	1	0	3	0	2	0	1	1	0	0	1	1	0	2/8
10. Nadya GR	0	2	2	0	1	0	0	0	1	0	2	0	1	1	7/3
11. Slava MOT	2	0	2	0	2	0	2	0	2	0	0	0	0	2	10/2
12. Lera MOT	2	0	0	2	2	1	0	1	0	2	0	2	0	2	4/10
13. Tolya GR	0	1	0	2	0	3	1	0	0	2	0	2	0	2	1/12
14. Vova MOT	0	2	0	2	0	3	0	1	3	0	0	1	0	2	3/11
15. Inna MOT	0	2	0	3	0	3	0	1	1	1	0	1	0	2	1/13
16. Galya MOT	0	2	1	1	0	1	0	1	1	1	0	2	0	2	2/10
17. Vera MOT	0	2	0	2	0	2	0	1	0	2	0	0	0	2	0/11
18. Liza MOT	0	2	0	2	0	1	0	1	0	2	0	0	0	0	0/8
19. Denis FAT	0	2	0	2	0	1	0	0	0	2	0	2	0	1	0/10
20. Polyu GR	0	1	0	1	0	1	1	0	0	1	0	2	0	2	1/8
21. Sonya SIS	0	1	0	1	0	1	0	1	3	1	0	1	1	1	4/7
Total	4	29	8	36	7	38	6	16	20	17	5	28	3	29	53/193

Table 10.15: Agreement production with individual hybrid nouns referring to females (12 children, age 5;1-6;5)

Child	Age	<i>počítal on</i> 'postwoman'		<i>doktor</i> 'doctor'		<i>milicioner</i> 'policewoman'		<i>fotograf</i> 'photographer'		<i>povar</i> 'cook'		<i>vráč</i> 'physician'		<i>bibliotekar'</i> 'librarian'		Total M/F
		M	F	M	F	M	F	M	F	M	F	M	F			
1. Borya	5;1	2	0	2	0	2	0	1	0	0	0	0	0	0	0	7/0
2. Valya	5;1	1	1	3	0	2	0	1	0	2	0	1	0	1	0	11/1
3. Vitya	5;3	2	0	2	0	2	0	1	0	2	0	1	0	2	0	12/0
4. Paša	5;3	2	0	2	0	2	0	1	0	2	0	2	0	1	1	12/1
5. Syeta	5;4	1	1	1	1	0	2	1	0	1	1	1	1	2	0	7/6
6. Nina	5;7	2	0	2	0	2	0	1	0	2	0	1	1	1	1	11/2
7. Anya	5;8	2	0	2	0	2	0	1	0	0	0	0	0	0	0	7/0
8. Stas	5;9	1	1	1	0	1	0	0	1	1	0	0	2	0	2	4/6
9. Jana	5;11	1	1	1	0	3	0	1	0	1	0	1	0	0	2	8/3
10. Yulya	5;1	1	0	0	2	0	2	1	0	2	0	0	2	0	2	4/8
11. Tanya	6;3	2	0	1	1	0	3	1	0	2	0	0	1	0	2	6/7
12. Dusya	6;5	2	0	1	0	2	0	1	0	2	0	2	0	0	2	10/2
Total		19	4	18	4	18	7	11	1	17	1	9	7	7	12	99/36

Table 10.16: Agreement production with hybrid nouns, referent unspecified
(25 children, aged 2;6-4;0)

Child	Age	M	F	Total
1. Dima	2;6	1	6	7
2. Olya	2;7	9	0	9
3. Petya	2;8	18	0	18
4. Roma	2;10	19	0	19
5. Katya	2;11	14	0	14
6. Seva	3;0	14	0	14
7. Lena	3;1	16	0	16
8. Kolya	3;1	2	9	11
9. Ira	3;2	23	0	23
10. Vasya	3;3	11	0	11
11. Lyuba	3;3	15	1	16
12. Nadya	3;3	13	0	13
13. Slava	3;6	9	0	9
14. Lera	3;6	16	0	16
15. Tolya	3;6	11	0	11
16. Vova	3;7	16	0	16
17. Inna	3;9	21	0	21
18. Galya	3;9	17	0	17
19. Vera	3;9	18	0	18
20. Liza	3;9	18	0	18
21. Denis	3;9	16	0	16
22. Polya	3;10	19	0	19
23. Sonya	3;10	14	0	14
24. Roza	3;10	18	0	18
25. Oleg	4;0	18	0	18
Total		366	16	382

Table 10.17: Agreement production with double gender nouns, referent unspecified (25 children, age 2;6-4;0)

Child	Age	<i>narjažalka</i> 'decorator'		<i>žalalka</i> 'pitiful man'		<i>umnjaša</i> 'smarty pants'		<i>pomogaša</i> 'helper'		<i>pačikara</i> 'sloven'		Total M/F/PL
		M	F	M	F	M	F	M	F	M	F	
1. Dima	2;6	0	0	0	0	0	1	0	2	0	2	0/5/0
2. Olya	2;7	1pl	1	0	2	2	0	3	0	2	0	7/3/1
3. Petya	2;8	1pl	0	0	3	3	0	3	0	0	2	6/5/1
4. Roma	2;10	2pl	0	0	2	5	0	4	0	2	0	11/2/2
5. Katya	2;11	0	2	0	2	2	1	0	3	0	2	2/10/0
6. Seva	3;0	1pl	1	1pl	2	3	0	3	0	1	0	7/3/2
7. Lena	3;1	0	2	1	2	3	0	0	4	1	0	5/8/0
8. Kolya	3;1	0	0	1	0	2	0	2	0	0	0	5/0/0
9. Ira	3;2	2pl	0	1	1	2	0	2	0	2	0	7/1/2
10. Vasya	3;3	1	0	2	0	1	0	2	0	2	0	8/0/0
11. Lyuba	3;3	0	0	0	2	1	0	4	0	2	0	7/2/0
12. Nadya	3;3	0	1	0	2	1pl	1	0	2	0	2	0/8/1
13. Slava	3;6	1	0	2	0	2	0	2	1	2	0	9/1/0
14. Lera	3;6	0	2	1	1	2	1	0	3	0	3	3/10/0
15. Tolya	3;6	3pl	0	3pl	0	2pl	0	3pl	0	2pl	0	0/0/13
16. Vova	3;7	2pl	1	0	4	3	0	3	0	2	0	8/5/2
17. Inna	3;9	0	1	3	0	3	0	3	0	2	0	11/1/0
18. Galya	3;9	0	2	0	3	0	2	0	3	0	2	0/12/0
19. Vera	3;9	1pl	0	2	0	3	0	2	0	1	0	8/0/1
20. Liza	3;9	4	0	3	0	4	0	3	0	1	2	15/2/0
21. Denis	3;9	1pl	1	2	0	2	1pl	4	0	2	0	10/1/2
22. Polya	3;10	0	2	4	0	2	0	3	0	0	2	9/4/0
23. Sonya	3;10	1	1	0	3	2	0	0	3	1	0	4/7/0
24. Roza	3;10	0	2	0	2	0	3	0	3	0	2	0/12/0
25. Oleg	4;0	2	0	1pl	2	3	0	1	0	2	0	8/2/1
Total												150/104/28

Table 10.18: Agreement production with double gender nouns referring to males (25 children, age 2;6-4;0)

Child	Age	<i>umnjaša</i> 'smarty pants'		<i>stijaga</i> 'mod'		<i>pačkava</i> 'sloven'		<i>obižala</i> 'bully'		<i>poedala</i> 'heavy eater'		<i>bedolaga</i> 'poor wretch'		Total M/F
		M	F	M	F	M	F	M	F	M	F	M	F	
1. Dima	2;6	1	0	1	0	0	0	0	0	4	0	1	0	7/0
2. Olya	2;7	1	0	4	0	2	0	2	0	2	0	2	0	13/0
3. Petya	2;8	3	0	5	0	3	0	2	0	2	0	3	0	18/0
4. Roma	2;10	3	0	3	1	2	0	2	0	2	0	4	0	16/1
5. Katya	2;11	1	2	2	2	0	3	2	0	1	2	3	1	9/10
6. Seva	3;0	2	0	2	1	3	0	2	0	2	0	4	0	15/1
7. Lena	3;1	3	0	6	0	2	0	1	0	3	0	4	0	19/0
8. Kolya	3;1	2	1	3	1	1	0	0	0	1	0	1	2	8/4
9. Ira	3;2	1	1	3	0	3	0	2	0	2	0	5	0	16/1
10. Vasya	3;3	3	0	2	0	1	0	1	0	2	0	2	0	11/0
11. Lyuba	3;3	3	0	4	0	3	0	2	0	1	0	3	0	16/0
12. Nadya	3;3	3	0	4	0	3	0	2	0	1	2	2	0	15/2
13. Slava	3;6	2	0	3	0	2	0	1	0	0	2	2	1	10/3
14. Lera	3;6	4	0	2	0	0	2	2	0	2	2	3	0	13/4
15. Tolya	3;6	1	0	2	0	2	0	2	0	2	0	2	0	11/0
16. Vova	3;7	2	0	5	0	1	0	2	0	2	0	2	0	14/0
17. Inna	3;9	3	0	5	0	2	0	3	0	6	0	3	0	22/0
18. Galya	3;9	3	0	1	2	1	1	1	0	1	0	2	0	9/3
19. Vera	3;9	2	0	3	0	2	0	2	0	1	1	3	0	13/1
20. Liza	3;9	2	1	3	0	2	0	2	0	1	0	4	0	14/1
21. Denis	3;9	3	0	3	0	3	0	1	0	2	0	3	0	15/0
22. Polyu	3;10	3	0	4	0	3	0	2	0	1	0	4	0	17/0
23. Sonya	3;10	3	0	4	0	3	0	1	0	4	0	3	0	18/0
24. Roza	3;10	3	0	2	0	2	0	2	0	2	0	2	0	13/0
25. Oleg	4;0	3	0	3	0	2	0	2	0	3	0	4	1	17/1
Total		60	5	79	7	48	6	41	0	50	9	71	5	349/32

Table 10.19: Agreement production with double gender nouns referring to females (25 children, age 2;6-4;0)

Child	Age	'flaky person' <i>terjaca</i>		'smarty pants' <i>umnjaša</i>		'pitiful man' <i>žalečka</i>		'sloven' <i>pačkara</i>		Total M/F
		M	F	M	F	M	F	M	F	
1. Dima	2;6	0	0	0	2	0	1	0	1	0/4
2. Olya	2;7	0	2	0	0	0	2	2	0	2/4
3. Petya	2;8	0	2	3	0	0	2	0	3	3/7
4. Roma	2;10	0	2	1	3	0	1	0	1	1/7
5. Katya	2;11	0	3	0	3	0	3	0	4	0/13
6. Seva	3;0	0	3	0	3	0	1	0	1	0/8
7. Lena	3;1	0	4	0	3	0	3	0	3	0/13
8. Kolya	3;1	0	2	1	2	0	2	0	1	1/7
9. Ira	3;2	0	3	0	3	0	2	0	0	0/8
10. Vasya	3;3	2	0	4	0	2	0	2	0	10/0
11. Lyuba	3;3	0	2	0	3	0	3	0	3	0/11
12. Nadya	3;3	0	3	0	2	0	3	0	1	0/9
13. Slava	3;6	1	0	0	3	0	1	0	2	1/6
14. Lera	3;6	0	2	0	3	0	3	0	3	0/11
15. Tolya	3;6	0	1	0	0	0	0	0	1	0/2
16. Vova	3;7	0	2	2	0	0	2	2	1	4/5
17. Imma	3;9	0	2	0	3	0	2	0	3	0/10
18. Galya	3;9	0	2	0	2	0	2	0	1	0/7
19. Vera	3;9	0	2	1	2	0	2	0	3	1/9
20. Liza	3;9	0	2	0	3	0	3	0	4	0/12
21. Denis	3;9	0	2	0	3	0	3	0	3	0/11
22. Polya	3;10	0	3	0	2	0	2	0	3	0/10
23. Sonya	3;10	0	1	0	2	0	2	2	1	2/6
24. Roza	3;10	0	3	0	3	0	2	0	1	0/9
25. Oleg	4;0	0	2	3	0	0	1	0	2	3/5
Total		3	50	15	50	2	48	8	46	28/194

Appendix III

Figure 10.1: Agreement distribution with double gender nouns used with reference to males and females across three age groups (24 children, age 2;6-4;0)

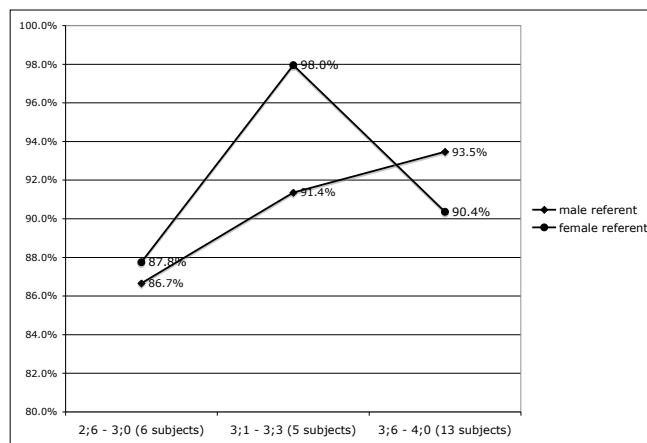
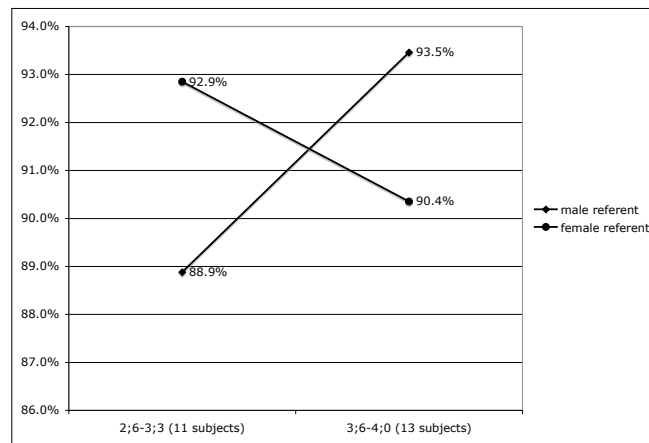


Figure 10.2: Agreement distribution with double gender nouns used with reference to males and females across two age groups (24 children, age 2;6-4;0)



Bibliography

- Alexiadou, A. 2004. Inflection class, gender and DP internal structure. In *Explorations in Nominal Inflection*, ed. A. Alexiadou and T. A. Hall, 21–49. Berlin: Mouton de Gruyter.
- Aronoff, M. 1994. *Morphology by Itself: Stems and Inflectional Classes*. Cambridge, Mass.: The MIT Press.
- Aronoff, M., and K. Fudeman. 2005. *What is Morphology?*. Malden, Mass.: Blackwell.
- Baauw, S. 2001. Expletive determiners in child Dutch and Spanish. In *BUCLD Proceedings 25*, ed. A.H.-J. Do et al, 1–12. Somerville, Mass.: Cascadilla Press.
- Baauw, S. 2002. *Grammatical Features and the Acquisition of Reference: A Comparative Study of Dutch and Spanish*. New York and London: Routledge.
- Barbier, I. 2000. An experimental study of scrambling and object shift in the acquisition of Dutch. In *The Acquisition of Scrambling and Cliticization*, ed. S. M. Powers and C. Hamann, 41–69. Dordrecht: Kluwer Academic Publishers.
- Berko Gleason, J. 1958. The child's learning of English morphology. *Word* 14:150–177.
- Berman, R. A. 1985. The acquisition of Hebrew. In *The Crosslinguistic Study of Language Acquisition*, ed. D. I. Slobin, 255–371. NJ: Lawrence Erlbaum: Hillsdale.
- Bloom, P. 1990. Syntactic distinctions in child language. *Journal of Child Language* 17:343–355.
- Bloom, P. 2000. *How Children Learn the Meanings of Words*. Cambridge, Mass.: MIT Press.

- Brown, R. 1973. *A First Language*. Cambridge, Mass.: Harvard University Press.
- Brown, R., and C. Hanlon. 1970. Derivational complexity and order of acquisition in child speech. In *Cognition and the Development of Language*, ed. J. Hayes, 11–53. New York: Wiley.
- Burge, T. 1973. Reference and proper names. *The Journal of Philosophy* LXX:425–439.
- Chomsky, N. 1959. Review of B. F. Skinner's Verbal Behavior. *Language* 35:26–58.
- Chomsky, N. 1965. *The Logical Structure of Linguistic Theory*. New York: Plenum Press.
- Chomsky, N. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.
- Chomsky, N. 1986. *Knowledge of Language: Its Nature, Origin, and Use*. London: Praeger.
- Clark, E. V. 1985. The acquisition of Romance with a special reference to French. In *The Crosslinguistic Study of Language Acquisition*, ed. D. I. Slobin, 687–782. Hillsdale, NJ: Lawrence Erlbaum.
- Corbett, G. G. 1982. Gender in Russian: An account of gender specification and its relationship to declension. *Russian Linguistics* 6:197–232.
- Corbett, G. G. 1983. *Hierarchies, Targets and Controllers: Agreement Patterns in Slavic*. London: Croom Helm.
- Corbett, G. G. 1991. *Gender*. Cambridge: Cambridge University Press.
- Corbett, G. G. 2006. *Agreement*. Cambridge: Cambridge University Press.
- Corbett, G. G., and N. M. Fraser. 2000. Gender assignment: A typology and a model. In *Systems of Nominal Classification*, ed. G. Senft, 293–325. Cambridge: Cambridge University Press.
- Crain, S., and R. Thornton. 1998. *Investigations in Universal Grammar: A Guide to Experiments on the Acquisition of Syntax and Semantics*. Cambridge, Mass.: MIT Press.
- Crain, S., and K. Wexler. 1999. Methodology in the study of language acquisition: A modular approach. In *Handbook of Child Language Acquisition*, ed. W. C. Ritchie and T. K. Bhatia, 387–425. San Diego: Academic Press.

- Crockett, D. B. 1976. *Agreement in Contemporary Russian*. Cambridge, Mass.: Slavika Publishers, Inc.
- Dahl, Ö. 2000. Animacy and the notion of semantic gender. In *Gender in Grammar and Cognition*, ed. B. Unterbeck, 99–115. Berlin: Mouton de Gruyter.
- De Houwer, A. 1990. *The Acquisition of Two Languages from Birth: A Case Study*. Cambridge: Cambridge University Press.
- Demetras, M. J. et al. 1986. Feedback to first language learners: The role of repetitions and clarification questions. *Journal of Child Language* 13:275–292.
- Enger, H. O. 2004. On the relation between gender and declension: A diachronic perspective from Norwegian. *Studies in Language* 28:51–82.
- Gallaway, C., and B. J. Richards, ed. 1994. *Input and Interaction in Language Acquisition*. Cambridge: Cambridge University Press.
- Gentner, D. 1982. Why nouns are learnt before verbs: Linguistic relativity versus natural partitioning. In *Language Development*, ed. Kuczaj S. A., volume 2, 301–334. NJ: Lawrence Erlbaum: Hillsdale.
- Gordishevsky, G., and S. Avrutin. 2004. Optional omission in an optionally null subject language. In *Proceedings of GALA 2003*, ed. J. van Kampen and S. Baaui, 187–198. Utrecht: LOT.
- Gould, J. L., and P. Marler. 1987. Learning by instinct. *Scientific American* 255:74–85.
- Graudina, L. K. 1976. *Grammatičeskaja Pravil'nost' russkoj Reči. Opyt Častotno-stilističeskogo Slovarja Variantov*. Moscow: Nauka.
- Graudina, L. K. 1977. Slovoizmenitelnye varianty: Ves peremennyx elementov v grammatike (formy suščestvitel'nyx). In *Grammatika i Norma*, ed. L. K. Graudina et al, 144–170. Moscow: Nauka.
- Greenberg, J. H. 1978. How does a language acquire gender markers? In *Universals of Human Language*, ed. J. H. Greenberg et al, volume III, Word Structure, 42–82. Stanford, Ca.: Stanford University Press.
- Groebner, D. et al, ed. 2005. *Business Statistics: A Decision-Making Approach*. Upper Saddle River, NJ: Prentice Hall.

- Gvozdev, A. N. 1961. *Formirovanie u Rebenka Grammatičeskogo Stroja Russkogo Jazyka*. Moscow: APN RSFSR.
- Hall, D. G. 1996. Preschoolers' default assumptions about word meaning: Proper names designate unique individuals. *Developmental Psychology* 32:177–186.
- Hall, D. G., and S. A. Graham. 1997. Beyond mutual exclusivity: Children use form class information to constrain word-referent mapping. In *Proceedings of the Twenty-ninth Stanford Child Language Research Forum*, ed. E. Clark. New York: Cambridge University Press.
- Henzl, V. M. 1975. Acquisition of grammatical gender in Czech. *Papers and Reports on Child Language Development* 10:188–200.
- Hockett, C. F. 1958. *A Course in Modern Linguistics*. New York: Macmillan.
- Hyams, N. 1986. *Language Acquisition and the Theory of Parameters*. Dordrecht: Reidel.
- Iomdin, L. L. 1990. *Avtomatičeskaja Obrabotka Teksta na Estestvennom Jazyke: Model' Soglasovanija*. Moscow: Nauka.
- Josefsson, G. 1996. The acquisition of object shift in Swedish child language. In *Children's Language 9*, ed. G. H. V. Gilbert, 153–165. Mahwah: Lawrence Erlbaum.
- Jusczyk, P. 1997. *The Discovery of Spoken Language*. Cambridge, Mass.: MIT Press.
- Kapatsinski, V. 2006. Sex associations of Russian generics. *Journal of Slavic Linguistics* 14:17–43.
- Karmiloff-Smith, A. 1979. *A Functional Approach to Child Language: A Study of Determiners and Reference*. Cambridge: Cambridge University Press.
- Kasatkin, L. L., ed. 2005. *Russkaja Dialektologija*. Moscow: ACADEMIA.
- Kempe, V. 2003. Diminutivization supports gender acquisition in Russian. *Journal of Child Language* 30:471–485.
- Kerswill, P., and A. Williams. 2000. Creating a new town koine: Children and language change in Milton Keynes. *Language in Society* 29:65–115.

- Kiparsky, P. 1973. 'Elsewhere' in phonology. In *A Festschrift for Morris Halle*, ed. S. R. Anderson and P. Kiparsky, 93–106. New York: Holt, Rinehart and Winston.
- Kopeliovč, A. B. 1977. K voprosu o kodifikaciji imen suščestvitel'nyx obščego roda. In *Grammatika i Norma*, ed. L. K. Graudina et al, 178–192. Moscow: Nauka.
- Kupisch, T. 2002. Gender in monolingual and bilingual first language acquisition: Comparing Italian and French. *Lingue e Linguaggio* 1:107–149.
- Kupisch, T. 2007. Testing the effects of frequency on the rate of learning: Determiner use in early French, German and Italian. In *Frequency Effects in Language Acquisition. Studies on Language Acquisition SOLA series*, ed. I. Gülzow and N. Gagarina, 83–114. Berlin: Mouton De Gruyter.
- Kupisch, T. to appear. A fresh look at root infinitives from a cross-linguistic perspective. *Nordlyd* .
- Labov, W. 2001. *Principles of Linguistic Change. Volume 2: Social Factors*. Oxford: Blackwell.
- Levy, Y. 1983a. The acquisition of Hebrew plurals: The case of missing gender category. *Journal of Child Language* 10:107–121.
- Levy, Y. 1983b. It's frogs all the way down. *Cognition* 15:75–93.
- Lightfoot, D. W., ed. 1979. *Principles of Diachronic Syntax*. Cambridge: Cambridge University Press.
- Lightfoot, D. W. 1999. *The Development of Language: Acquisition, Change, and Evolution*. Malden, Mass. and Oxford: Blackwell.
- Lopatin, V. V. et al. 1989. *Sovremennyj Russkij Jazyk: Teoretičeskij Kurs: Slovoobrazovanie, Morfologija*. Moscow: Russkij jazyk.
- Lust, B. 2006. *Child Language Acquisition and Growth*. Cambridge: University Press.
- Macnamara, J. 1982. *Names for Things*. Cambridge, Mass.: MIT Press.
- MacWhinney, B. 2000. *The CHILDES Project: Tools for Analyzing Talk*. Mahwah, N.J.: Lawrence Erlbaum.

- Maratsos, M. 1982. The child's construction of grammatical categories. In *Language Acquisition: The State of Art*, ed. E. Wanner and L. R. Gleitman, 240–266. Cambridge: Cambridge University Press.
- Marcus, G. F. 1993. Negative evidence in language acquisition. *Cognition* 46:53–85.
- Marcus, G. F. et al. 1992. *Overregularization in Language Acquisition*. Chicago: University of Chicago Press.
- Martynyuk, A. 1990. A contrastive study of male and female occupation terms in English and Russian. *Papers and Studies in Contrastive Linguistics* 26:103–110.
- McWhorter, J. 2006. Inflectional morphology and Universal Grammar: *Post hoc* versus *propter hoc*. Lecture presented at the PhD course 'Language contact from a political and social perspective', University of Tromsø, November 6–11, 2006.
- Mills, A. E. 1986. *The Acquisition of Gender: A Study of English and German*. Berlin: Springer.
- Mulford, R. 1985. Comprehension of Icelandic pronoun gender: Semantic versus formal factors. *Journal of Child Language* 12:443–453.
- Müller, N. 1994. Gender and number agreement within DP. In *Bilingual First Language Acquisition: French and German Grammatical Development*, ed. J. M. Meisel, 53–88. Amsterdam: J. Benjamin.
- Müller, N. 2000. Gender and number acquisition. In *Gender in Grammar and Cognition*, ed. B. Unterbeck, 351–399. Berlin: Mouton de Gruyter.
- Mučnik, I. P. 1971. *Grammatičeskie Kategorii Glagola i Imeni v Sovremenom Russkom Literaturnom Jazyke*. Moscow: Nauka.
- Nesset, T. 2003. Gender assignment in Ukrainian: Language specific rules and universal principles. *Poljarnyj Vestnik* 6:71–85.
- Nesset, T. 2006. Gender meets the usage-based model: Four principles of the rule interaction in gender assignment. *Lingua* 116:1369–1393.
- Nikunlassi, A. 2000. On gender assignment in Russian. In *Gender in Grammar and Cognition*, ed. B. Unterbeck, 771–791. Berlin: Mouton de Gruyter.

- Panov, M. V. 1968. *Morfologija i Sintaksis Sovremennogo Russkogo Literaturnogo Jazyka (Russkij Jazyk i Sovetskoe Obscestvo : Sociolinguisticeskoe Issledovanie)*. Moscow: Nauka.
- Pinker, S. 1982. A theory of the acquisition of lexical interpretive grammars. In *The Mental Representation of Grammatical Relations*, ed. J. Bresnan, 655–726. Cambridge, Mass.: MIT Press.
- Pinker, S. 1984. *Language Learnability and Language Development*. Cambridge, Mass.: Harvard University Press.
- Pinker, S. 1995. Why the child holds the baby rabbit: A case study in language acquisition. In *An Invitation to Cognitive Science: Language*, ed. L. Gleitman and M. Liberman, 107–133. Cambridge, Mass.: MIT Press.
- Pinker, S. 1999. *Words and Rules: The Ingredients of Language*. New York: Basic Books.
- Polinsky, M. 1995. Cross-linguistic parallels in language loss. *Southwestern Journal of Linguistics* 14:87–123.
- Polinsky, M. in press. Gender under incomplete acquisition: Heritage speakers' knowledge of noun categorization. Available at <http://ling.ucsd.edu/polinsky/publications.html> .
- Popova, M. I. 1973. Grammatical elements of language in the speech of pre-school children. In *Studies of Child Language Development*, ed. C. A. Ferguson and D. I. Slobin, 269–280. New York: Holt, Rinehart & Winston.
- Rice, C. 2006. Optimizing gender. *Lingua* 116:1394–1417.
- Rizzi, L. 1982. *Issues in Italian Syntax*. Dordrecht: Foris.
- Roberts, J. 1997a. Acquisition of variable rules: A study of (-t,d) deletion in preschool children. *Journal of Child Language* 24:351–372.
- Roberts, J. 1997b. Hitting a moving target: Acquisition of sound changes by Philadelphia children. *Language Variation and Change* 9:249–266.
- Roberts, J. 2002. Child language variation. In *The Handbook of Language Variation and Change*, ed. J. K. Chambers et al, 333–348. Oxford: Blackwell.

- Roeper, T. 2007. What frequency can do and what it can't. Frequency effects in language acquisition: Defining the limits of frequency as an explanatory concept. In *Frequency Effects in Language Acquisition. Studies on Language Acquisition SOLA Series*, ed. I. Gülzow and N. Gagarina, 23–48. Berlin: Mouton De Gruyter.
- Rumelhart, D., and J. McClelland. 1986. On learning the past tenses of English verbs: Implicit rules or parallel distributed processing? In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, ed. D. Rumelhart J. McClelland and PDP Research Group, 216–271. Cambridge, Mass.: MIT Press.
- Schaeffer, J. C. 2000. *The Acquisition of Direct Object Scrambling and Clitic Placement: Syntax and Pragmatics*. Amsterdam: Benjamins.
- Slobin, D. I. 1973. Cognitive prerequisites for the development of grammar. In *Studies of Child Language Development*, ed. C. A. Ferguson and D. I. Slobin, 175–208. New York: Holt, Rinehart and Winston.
- Slobin, D. I. 1985. Crosslinguistic evidence for the language-making capacity. In *The Crosslinguistic Study of Language Acquisition*, ed. D. I. Slobin, 1157–1256. Hillsdale, NJ: Lawrence Erlbaum.
- Smith, J. et al. 2007. “Mam, my trousers is fain doon!”: Community, caregiver, and child in the acquisition of variation in a Scottish dialect. *Language Variation and Change* 19:63–99.
- Smoczyńska, P. M. 1985. The acquisition of Polish. In *The Crosslinguistic Study of Language Acquisition*, ed. D. I. Slobin, 595–686. NJ: Lawrence Erlbaum: Hillsdale.
- Steinmetz, D. 1985. Gender in German and Icelandic: Inanimate nouns. In *Germanic linguistics. Papers from a symposium at the University of Chicago*, ed. J. T. Faarlund, 10–28. Bloomington, IN: IULC.
- Steinmetz, D. 1986. Two principles and some rules for gender in German: Inanimate nouns. *Word* 37:189–217.
- Švedova, N. Ju. et al. 1980. *Russkaja Grammatika*. Moscow: Nauka.
- Svenonius, P. 2007. Interpreting uninterpretable features. *Linguistic Analysis* 33.3-4:375–413.
- Theakston, A. 2004. Semantic generality, input frequency and the acquisition of syntax. *Journal of Child Language* 31:61–99.

- Thornton, R. 1996. Elicited production. In *Methods for Assessing Children's Syntax*, ed. D. McDaniel et al, 55–76. Cambridge, Mass.: MIT Press.
- Tomasello, M. 2000. Do young children have adult syntactic competence? *Cognition* 74:209–253.
- Tomasello, M. 2003. *Constructing a Language: A Usage-based Model of Language Acquisition*. Cambridge, Mass.: Harvard University Press.
- Čukovskij, K. 1965. *Sobranie Sočinenij v Šesti Tomax*, volume 1. Khudožestvennaja literatura.
- Voeykova, M. D. 1997. Acquisition of adjectival inflections (secondary paradigms in child Russian). *Papers and Studies in Contrastive Linguistics* 33:141–151.
- Westergaard, M. R. 2006. Triggering V-2: The amount of input needed for parameter setting in a Split-CP model of word order. In *Proceedings of GALA 2005*, ed. A. Belletti et al, 658–671. Cambridge: Cambridge Scholars Press.
- Westergaard, M. R. 2007. Word order in subject and object shift constructions in early child language: Pragmatics or economy? Guest lecture at the University of Lund.
- Westergaard, M. R. forthcoming. Verb movement and subject placement in the acquisition of word order: Pragmatics or structural economy? In *First Language Acquisition of Morphology and Syntax: Perspectives Across Languages and Learners*, ed. P. Guijarro-Fuentes et al, Language Acquisition and Language Disorders. Amsterdam and Philadelphia: Benjamins.
- Westergaard, M. R., and K. Bentzen. 2007. The (non-)effect of input frequency on the acquisition of word order in Norwegian embedded clauses. In *Frequency Effects in Language Acquisition. Studies on Language Acquisition SOLA Series*, ed. I. Gülzow and N. Gagarina, 271–306. Berlin: Mouton De Gruyter.
- Wexler, K. 1998. Very early parameter setting and the unique checking constraint: A new explanation of the optional infinitive stage. *Lingua* 106:23–79.
- Yang, C. D. 2000. Internal and external forces in language change. *Language Variation and Change* 12:231–250.

- Yang, C. D. 2002. *Knowledge and Learning in Natural Language*. New York: Oxford University Press.
- Zakharova, A. V. 1973. Acquisition of forms of grammatical case by preschool children. In *Studies of Child Language Development*, ed. C. A. Ferguson and D. I. Slobin, 281–284. New York: Holt, Rinehart & Winston.
- Zdorenko, T. 2005. The Acquisition of Subjects in English and Russian. Master's thesis, University of Tromsø.
- Zemskaja, E. A. et al. 1981. *Russkaja Razgovornaja Reč*. Moscow: Nauka.